GRADE RETENTION AND THE AFFECTIVE COMPONENTS OF LEARNING: EFFECTS ON STUDENTS' SELF-CONCEPT, SELF-ESTEEM, GOAL ORIENTATIONS, AND SCHOOL BELONGING

Joana Pipa

Thesis supervised by

Francisco Peixoto, Ph.D.

CIE-Ispa, Ispa-Instituto Universitário

Tese submetida como requisito parcial para obtenção do grau de

Doutoramento em Psicologia

Área de especialidade ..... Psicologia Educacional

2023

2023

Tese apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Psicologia na área de especialização de Psicologia Educacional realizada sob a orientação de Francisco Peixoto, apresentada no Ispa - Instituto Universitário no ano de 2023.

O presente trabalho foi financiado pela Fundação para a Ciência e Tecnologia (SFRH/BD/132195/2017).

To JoãoB, JoãoD, and JoãoP.

To Sebastião, the most incredible human being.

ACKNOWLEDGMENTS

“It takes a village” to raise a PhD. It’s been more than ten years (who’s counting?) since I started to follow this dream. My village helped me to travel such a long way and to finally cross the finish line. I will address them in Portuguese, otherwise I would get lost in translation: Obrigada.

Ao meu orientador, Prof. Doutor Francisco Peixoto, pela viagem que fizemos e que começou bem antes deste doutoramento. Por ter confiado em mim, sempre. Por não ter desistido quando tudo indicava que iria ser impossível ter uma bolsa. Pela disponibilidade, pelo conhecimento partilhado, por apaziguar as minhas inseguranças e, acima de tudo, por me deixar voar. Pelas oportunidades que me proporcionou e que tanto me fizeram crescer. O meu obrigada vai para além deste trabalho.

A todos os alunos, professores e escolas que, apesar das inúmeras tarefas e solicitações, generosamente nos abrem a porta e partilham connosco o seu tempo que já é tão pouco. Sem eles, nenhum dos trabalhos aqui apresentado existiria.

À Fundação para a Ciência e Tecnologia por ter financiado este trabalho.

Aos Professores de Psicologia Educacional e membros do CIE-Ispa, em especial à Professora Doutora Margarida Alves Martins, com quem muito cresci e aprendi sobre, entre muitas outras coisas, o lado B da investigação.

Aos meus colegas de seminário de doutoramento, obrigada pelas partilhas, entreajuda e acima de tudo, por contrariarem a solidão que um doutoramento muitas vezes traz.

À “malta do sótão” e restantes amigos, família que o ISPA me deu. A vontade de ir não seria a mesma se não vos encontrasse lá. E seguramente também não seria a mesma a vontade de ficar. Obrigada, Cris, Daniela, Eva, Joana, Marta, Nalini, Patrícia, Rita e Sérgio. Em especial, obrigada Cristina e Natalie por serem fonte de inspiração e por todo o apoio, discussão, partilhas e desabafos.

Ao João Daniel. Por saber tanto e por ser sempre tão generoso nessa partilha.

Aos meus amigos ‘não ispianos’, família que escolhi. Por darem cor à vida. Por celebrarem comigo, e por vezes até mais do que eu, as minhas conquistas.

Ao consórcio sobre retenção e à equipa do Projecto AlteR, com quem muito aprendi sobre retenção escolar e cujos trabalhos inspiraram este que aqui apresento. Em especial, agradeço à Mieke Goos. A nossa colaboração foi muito importante para este trabalho. Obrigada pela motivação e entusiamo contagiante, pela partilha de conhecimento e por acreditar em mim e neste trabalho.

À minha família, em especial aos meus pais e ao meu irmão, agradeço serem apoio e amparo. À minha nova família, em especial aos pais do João por cuidarem e por nas últimas semanas terem possibilitado que a minha jornada de trabalho ganhasse mais horas.

Ao João. Por ser o meu porto de abrigo e o meu companheiro. Por me ter incentivado desde o primeiro dia a embarcar nesta aventura. Por ter segurado o barco nos últimos tempos para que me pudesse dedicar a terminar este trabalho.

Por fim, dedico este trabalho ao meu filho, Sebastião. Ter nascido tua mãe foi a mais bonita das descobertas. Teres nascido meu filho foi a mais incrível das sortes. E está tudo certo.

Lisboa, 17 de Fevereiro de 2023

Palavras-chave:

Retenção escolar; autoconceito; motivação; identificação com a escola; meta-análise

Key-words:

Grade retention; self-concept, motivation; school belonging; meta-analysis

Categorias de Classificação da tese:

PsycINFO Classification Categories and Codes:

2800 Developmental Psychology

2840 Psychosocial & Personality Development

3500 Educational & School Psychology

3550 Academic Learning & Achievement

3560 Classroom Dynamics & Student Adjustment & Attitudes

ABSTRACT

Every year, many students need help to meet the academic standards and expectations required for the different academic subjects and grade levels. For these students, grade retention would be the most likely scenario at the final of the school year. Grade retention is one of the most debated educational measures, with researchers questioning for decades whether it would be an effective or detrimental practice for students. This ongoing debate results from the little consensus obtained from the empirical studies investigating grade retention effects on students' academic and non-academic outcomes, either in the short or long term. Particularly, mixed findings were obtained considering students' affective components of learning that constitute a valuable aspect of students' academic success and adjustment. The consequences of grade retention on this group of variables, such as students' self-concept, self-esteem, motivation, or engagement, have greatly concerned educators. Some advocate that grade retention would help students improve their self-beliefs and motivation because it would allow them more time to learn the academic content they were struggling with. Others claim that grade retention is a stressful, demotivating, and stigmatising event that would undermine students' affective experiences. Thus, a clear answer regarding the effectiveness of grade retention is still needed through more clarity provided by research investigating the effects of grade retention on students' academic and non-academic outcomes. In this dissertation, we aimed to contribute to that answer. For that purpose, we conducted a meta-analysis summarising the findings of studies conducted during the last two decades the on the effects of grade retention on students' achievement, psychosocial variables, academic career, and professional career. In addition, given the inconclusive findings regarding the impact of grade retention on students' affective components of learning, our second aim of this dissertation was to investigate the effects of grade retention on students' self-concept, self-esteem, goal orientations, sense of school belonging and sense of school valuing. For this purpose, we conducted a cross-sectional study using the Portuguese data from PISA 2018 assessment and a longitudinal study following students initially in the 7th grade for 3 to six years. Finally, we aimed to investigate the moderating effects of school retention composition on the relationship between grade retention and those affective components, considering the PISA 2018 data. The rationale behind this final aim was that students' experiences are shaped within the context, in this case, the school, where they pertain. Thus, grade retention experiences may differ whether students are retained in a school with more or few students in the same circumstances. Our findings provided little support for the effectiveness of grade retention, given that the negligible or adverse effects on students' outcomes outweigh the small benefits that have emerged, particularly on students' affective components of learning and in the long term. Moreover, our results revealed spillover effects of grade retention, affecting non-retained students. Based on the findings of this dissertation, we advise against the use of grade retention or, at least, caution in its application.

RESUMO

Todos os anos, um número considerável de alunos apresenta dificuldades em cumprir com as expectativas e metas académicas estabelecidas para as diferentes disciplinas e para o nível de ensino. Para estes alunos, a retenção escolar será o cenário mais provável no final do ano lectivo. A retenção escolar é uma das medidas educativas mais debatidas, sendo que os investigadores durante décadas os investigadores têm questionado sobre os seus benefícios ou malefícios para os alunos. Este debate em parte é alimentado pelos resultados menos consensuais dos estudos investigando os efeitos da retenção escolar em componentes académicas e não académicas e tanto a curto como a longo-prazo. Especificamente, considerando as componentes afectivas da aprendizagem que constituem um aspecto muito importante para o sucesso e ajustamento académicos, os resultados dos estudos têm apresentado alguma controvérsia. Ainda assim, as consequências da retenção em variáveis como o autoconceito, a auto-estima, a motivação ou o envolvimento dos alunos têm suscitado muitas preocupações por parte dos agentes educativos. Alguns argumentam que a retenção escolar poderá ajudar os alunos a sentirem-se mais confiantes e motivados, uma vez que lhes é concedido mais tempo para as aprendizagens escolares. Outros, afirmam que a retenção constitui um acontecimento bastante stressante, desmotivador e estigmatizante e que abala as experiências afectivas dos alunos. Desta forma, está ainda em aberto uma resposta precisa à questão da eficácia da retenção. Nesta dissertação pretendemos contribuir para essa resposta, através da realização de uma meta-análise dos estudos efectuados nas últimas duas décadas investigando os efeitos da retenção escolar no rendimento académico, ajustamento psicossocial, percurso académico e percurso profissional dos alunos. Adicionalmente, devido aos resultados inconclusivos relativos aos efeitos da retenção escolar nas componentes afectivas da aprendizagem, o nosso segundo objectivo foi o de investigar os efeitos da retenção escolar no autoconceito, auto-estima, orientações motivacionais, sentimento de pertença e sentimento de valorização da escola. Nesse sentido, realizámos um estudo transversal com recurso aos dados portugueses do PISA 2018 e um estudo longitudinal que acompanhou durante três a seis anos alunos inicialmente a frequentar o 7º ano de escolaridade. Finalmente, pretendemos investigar o efeito moderador das taxas de retenção ao nível da escola, utilizando os dados do PISA 2018. Este objectivo fundamentou-se na ideia de que o contexto, neste caso o contexto escolar, é determinante na formação das auto-representações e experiências académicas dos alunos. Neste caso, a experiência da retenção escolar poderá ser diferente consoante o aluno frequente uma escola com mais ou menos alunos nas mesmas circunstâncias. Os resultados desta dissertação contrariam a ideia de que a retenção escolar é uma prática eficaz, uma vez que os efeitos nulos ou negativos da retenção ultrapassam os pequenos benéficos encontrados, particularmente nas componentes afectivas da aprendizagem e a longo-prazo. Mais ainda, verificámos que a retenção não afecta apenas os alunos repetentes, mas também os seus pares a frequentar a mesma escola. Considerando estes resultados, recomendamos que sejam consideradas medidas alternativas à retenção, ou o seu uso moderado.

INDEX

TABLE INDEX

FIGURE INDEX

GENERAL INTRODUCTION

The main topic of this dissertation is one of the most debated educational measures: the practice of grade retention and, specifically, grade retention effects. Grade retention is the practice of holding students back in the same grade level, generally for low achievement reasons (Allen et al., 2009; Brophy, 2006; Jimerson, 2001). It is a widespread practice in many Education systems (Agasisti & Cordero, 2017; Dupriez et al., 2008; European Commission, 2020; Eurydice, 2011; Goos, Schreirer et al., 2013; Mons, 2007) and is regarded by many teachers and educators as an effective practice (Range et al., 2012; Santana, 2019; Santos et al., 2023; Tomchin & Impara, 1992; Young et al., 2019) that acts on students' best interests. Research on the effects of grade retention has a long tradition, as the first studies date back to the early twentieth century (Keyes, 1911, as cited in Jackson, 1975) and have been flourishing particularly during the last two decades (Allen et al., 2009; Valbuena et al., 2021). However, the findings from recent sound studies have questioned the effectiveness of this practice (e.g., Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009).

Generally, studies tend to evaluate the effects of grade retention on students' development of achievement (Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009). However, a growing body of research has claimed that grade retention has an impact on other variables, such as on students' affective components of learning, and other long-term variables, such as students' dropout and later professional career choices and opportunities (e.g., Jimerson, 2001; Jimerson et al., 2002; Ou & Reynolds, 2010; Valbuena et al., 2021; Xia & Kirby, 2009). Notably, studies aiming to provide more clarity regarding grade retention effectiveness, such as review studies, have directed little attention to research investigating the effects of grade retention on these psychosocial and long-term outcomes (Allen et al., 2009; Bright, 2011; Holmes, 1989; Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009). Particularly considering the psychosocial variables, namely the affective components of learning such as self-concept, motivation, or engagement (Goos, Van Damme, et al., 2013; Kretschmann et al., 2019; Lamote et al., 2014; Martin, 2011; Peixoto et al., 2016, 2017), previous literature reviews stated that studies investigating these outcomes besides being still scarce, have produced so far ambiguous and inconclusive findings (Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009). Moreover, the affective components of learning are related to academic achievement in a transactional and bi-directional way (Huang, 2011; Jimerson, 2001; Marsh, 2016; Peixoto et al., 2016), evidencing the importance of unravelling the effects of retention on these outcomes.

Grade retention has been identified in the literature as having detrimental consequences on students' self-esteem (Hwang & Cappella, 2018; Mathys et al., 2019; Martin, 2011), academic self-concept (Kretschmann et al., 2019; Peixoto et al., 2016), motivation

(Martin, 2011; Peixoto et al., 2016, 2017) and school belonging (Martin, 2011; Van Canegem et al., 2021). These results oppose several others claiming that grade retention brings benefits (or at least does not harm) to students' self-concept (Ehmke et al., 2010; Lamote et al., 2014; Marsh, 2016) and, to a lesser extent, positive effects on engagement (Wu et al., 2010). These contradicting findings illustrate the complexity of determining whether grade retention should be considered helpful or harmful. Notwithstanding this, the two opposing branches of studies recognise that grade retention affects students' perception of competence and self-worth and their affective experiences within the school context (Martin, 2011).

Regarding the role of the school context and its related environment, surprisingly, most studies have neglected the impact of such context in explaining grade retention effects, with few exceptions (Bietenbeck, 2014; Demanet & Van Houtte, 2013, 2016; Gottfried, 2013; Hong & Raudenbush, 2005; Hong & Yu, 2008; Marsh, 2016; Van Canegem et al., 2021). These studies stated that school retention composition, operationalised as the proportion of retained students in school, affects the whole school community. In particular, school retention composition is associated with higher misconduct (Demanet & Van Houtte, 2013), a lower number of same-grade peer relationships (Demanet & Van Houtte, 2016), and students' self-concept (Van Canegem et al., 2021). The limited number of studies addressing this topic and its findings suggest that more investigation is needed to unveil the impact of school retention composition on students' outcomes.

Grade retention rates have been found to fluctuate across schools and countries, reflecting differences in each Education system's organisation and regulations (Agasisti & Cordero, 2017; Goos, Schreier, et al., 2013; European Commission, 2020; Eurydice, 2011) but also within-country variations due to beliefs and attitudes towards retention (Bonvin et al., 2008; European Commission, 2020; Eurydice, 2011; Goos, Schreier, et al., 2013; Santana, 2019; Santos et al., 2023; Tomchin & Impara, 1992). Recent studies also found that the country's retention composition could dictate differences in students' psychosocial factors (Van Canegem et al., 2022a, 2022b). These considerations thus appeal to study grade retention considering its cross-country and within-country contexts.

Most notably, studies conducted in the Portuguese context still need to be made available. To the best of our knowledge, only a few studies considered the effects of grade retention on Portuguese students' achievement (Conboy, 2011; Conboy et al., 2013; Nunes et al., 2018; Pereira & Reis, 2014; Rosário et al., 2013), affective components of learning (Nascimento & Peixoto, 2012; Peixoto et al., 2016, 2017), and students' school career (Nunes et al., 2018; Santos et al., 2022). At this point, it should be mentioned that from an international perspective, Portuguese grade retention rates are among the highest (European Commission, 2020; Eurydice, 2011).

Finally, previous literature ascribed controversial findings in the outcomes of grade retention to the different age groups assessed. A great deal of studies focused on the effects of grade retention in kindergarten and primary education (e.g., Gleason et al., 2007; Goos, Van Damme, et al., 2013; Hwang & Cappella, 2018; Moser et al., 2012; Vandecandelaere, Vansteelandt, et al., 2016; Wu et al., 2010). However, being retained at the beginning of the school career may differ from repeating in more advanced grades (Wu et al., 2010). Being retained in more advanced grades could increase the effects of grade retention because the stigma associated with grade retention may be more prominent (Demanet & Van Houtte, 2016; García-Perez et al., 2014; Giano et al., 2022; Jimerson & Ferguson, 2007; Mathys et al., 2019). In addition, in Portugal, grade retention rates are higher among secondary school students (Conselho Nacional da Educação [CNE], 2020).

Hence, this dissertation aimed to investigate the effects of grade retention in Portuguese lower secondary education on the affective components of learning. Specifically, we aimed to provide empirical evidence on 1) the effectiveness of grade retention by summarising the effects of grade retention on several academic and non-academic outcomes and considering several moderating variables; and 2) the effects of grade retention within the Portuguese context on the affective components of learning, by analysing the effects of grade retention in regards to students' self-concept and self-esteem, motivational (goal) orientations, and sense school belonging and school valuing, either cross-sectionally and longitudinally.

Figure 1 presents an overview of the rationale and organisation of the present dissertation. Considering the aims of this dissertation, the current work is organised into five chapters. A reference list accompanies each chapter/section. In the first chapter, corresponding to the theoretical framework, we address the rationale that grounded this study, as well as elaborate on the state of the problem. In the second chapter, the aims of this dissertation are outlined.

The third chapter initiates the empirical section and presents a systematic review and a meta-analysis of 84 studies conducted between 2000 and 2019 investigating the effects of repeating a grade in primary or secondary education on both retained and non-retained students' academic achievement, psychosocial functioning, school career, and job career. In this study, we further addressed whether the effects of grade retention could differ according to a set of moderators: timing of retention, country; retention intervention; outcome domain; length of the follow-up; type of comparison, and method of the study. This study aimed to add to the previous reviews (Allen et al., 2009; Bright, 2011; Holmes, 1989; Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009) by addressing these variables and, thus, helping to add more clarity regarding the effectiveness of grade retention.

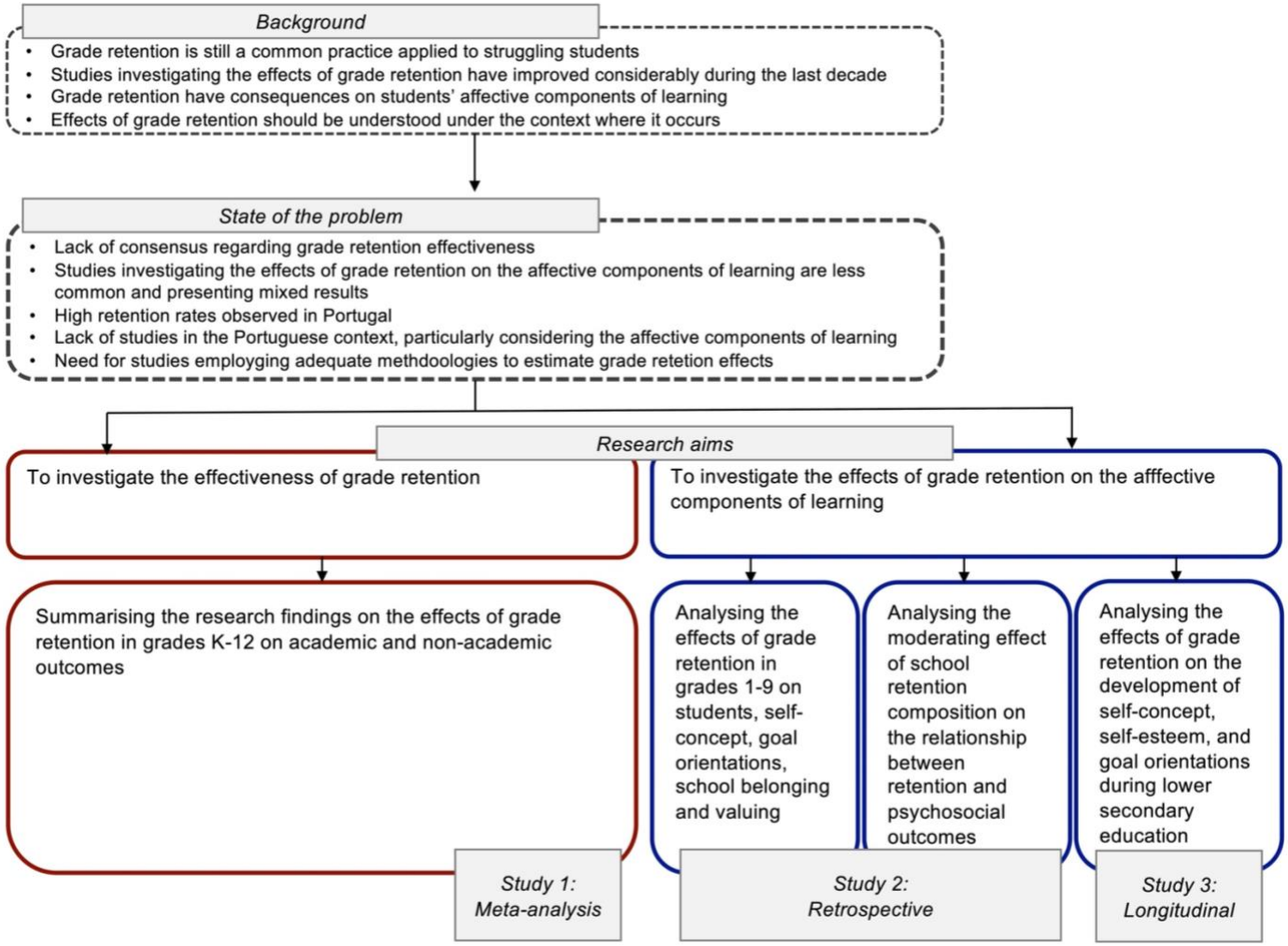
The fourth chapter presents an empirical study considering the Portuguese context. Through a cross-sectional and a retrospective study, we aimed to investigate the effects of grade retention occurring between grades 1 and 9 on students' reading self-concept, goal orientations, and sense of school belonging and valuing. Additionally, we explored whether the nature of these effects differed according to the proportion of retained students attending a school, i.e., school retention composition. To conduct this study, we used Portuguese data from the 2018 Programme for International Study Assessment (PISA) cycle. With this study, we aimed to shed some light on the effects of grade retention on students' affective components of learning, particularly considering the Portuguese context (Peixoto et al., 2016, 2017).

As for the fifth chapter, our goal was to provide more evidence regarding the effects of grade retention on the affective components of learning by relying on longitudinal research. Thus, based on the findings and suggestions from previous longitudinal studies (Hwang & Cappella, 2018; Klapproth et al., 2016; Kretschmann et al., 2019; Lamote et al., 2014; Vandecandelaere, Schmitt, et al., 2016), in this study, we investigated the effects of grade retention on the development of academic self-concept, self-esteem, and goal orientations during lower secondary education (i.e., grades 7 to 9) for continuously promoted, retained in 7th grade or retained in 8th grade students. In addition, based on a smaller sample, we conducted an exploratory analysis regarding the long-term effects of grade retention on the abovementioned variables and students' academic career.

The sixth and concluding chapter provides an integrative view of this dissertation's main findings and highlights the main contributions of this work. In addition, we address this study's limitations while suggesting future research investigating the impact of grade retention on the affective components of learning. Finally, to conclude this work, we reflect on the implications of our findings for educational policy and practice.

Figure 1

Overview of the Rationale, State of The Problem, and Aims of the Present Dissertation



SECTION I | THEORETICAL FRAMEWORK

1. THEORETICAL FRAMEWORK

1.1. Grade retention: One step back or one step ahead?

Every year, thousands of students struggle to meet the academic standards for a particular grade level. For teachers, dealing with students' heterogeneity in academic competencies and achievement is often seen as a challenge involving greater efforts to guarantee high educational standards. School systems have developed and applied different strategies to tackle students' heterogeneity (Dupriez et al., 2008), such as separating students according to different educational routes or tracks (popular among Germanic education systems) or relying on differentiated and individualised instruction (such as in the Scandinavian countries). Among other countries, for many struggling students repeating the grade level will be the most likely scenario. This dissertation focuses on this last strategy.

Grade retention, also referred to as grade repetition, flunking, nonpromotion, or being held back (Jimerson, 2001), is the practice of holding academically struggling students back in the same grade level for one additional year to give them more time to meet academic requirements for that grade level during the repeating year (Allen et al., 2009; Brophy, 2006; Jackson, 1975; Shepard & Smith, 1989). Under this definition, in this dissertation, we refer to retained students as those who remained in the same grade level for the whole school year instead of advancing to the next grade level along with their promoted peers – considered as nonretained or promoted students.

Defining grade retention could seem straightforward, however, it could take different forms. Grade retention could result from a voluntary decision initiated by students or their families or an involuntary-imposed practice initiated by the school (Brophy, 2006). Involuntary grade retention may be based on different measures: it could be a result of standardised achievement tests, or it could be a matter of teachers' and schools' judgments of students' academic situation (Bonvin et al., 2008; Brophy, 2006; Goos, Schreier et al., 2013; Xiang & Chiu, 2022). In addition, grade retention application also may differ across schools and countries. In some cases, retained students might be provided with additional remedial interventions such as tutoring, individualised teaching or summer schools (e.g., Allensworth, 2005; Brophy, 2006; Geng & Rockoff, 2017; Jimerson & Ferguson, 2007; Özek, 2015; Schwerdt et al., 2017), while in others grade retention comprises only a mere repetition of the subject curriculum (Brophy, 2006). The literature outlined in this work generally considers grade retention as an involuntary measure that typically involves mere repetition of subjects.

In contrast, grade promotion, social promotion, or automatic promotion is considered the practice of allowing students to advance to the next grade level with their class group, regardless of whether or not they have accomplished the academic standards for the grade level (Brophy, 2006; Thompson & Cunningham, 2000; Xia & Kirby, 2009). Social promotion is often criticised because it is perceived as a measure that sends unprepared students to the

next grade level, which will only increase their difficulties (Thompson & Cunningham, 2000; Xia & Kirby, 2009). Also, opponents of social promotion argue that it sends a message that a lack of work and effort is allowed and rewarded and that it mirrors a culture of educational facilitation and low academic standards (Brophy, 2006; Conboy et al., 2013; Thompson & Cunningham, 2000; Xia & Kirby, 2009).

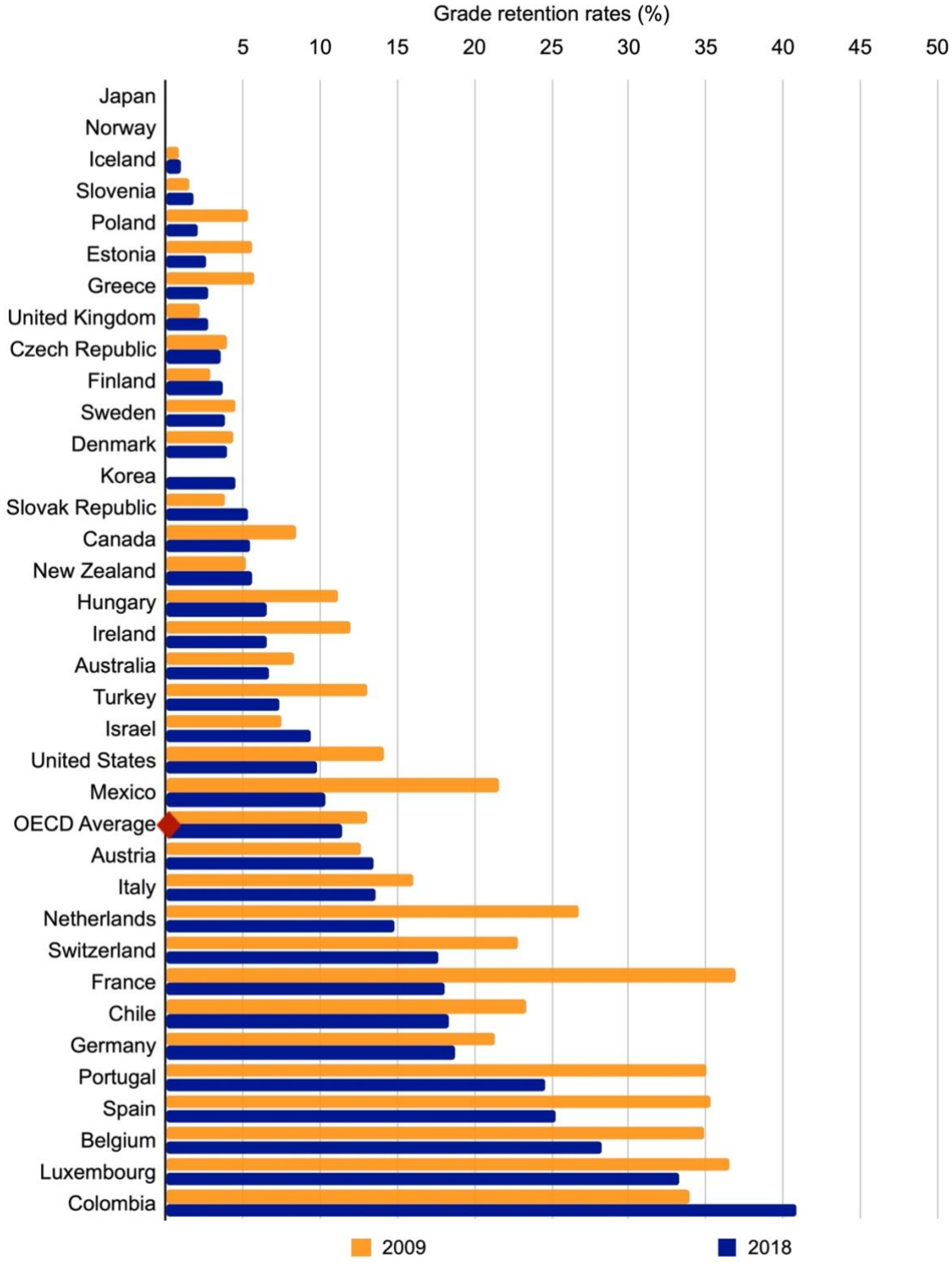
In practice, countries differ substantially in the application of grade retention, as can be noted in Figure 2, presenting the percentage of students of age 15 that reported having been retained at least once, derived from the PISA 2018 questionnaire (Organisation for Economic Co-operation and Development ([OECD], 2020). The recent data provided by the OECD (2020) show that grade retention is never applied in Norway and Japan and is seldomly applied in countries such as Norway, Japan, Iceland, Slovenia, Poland, Estonia, Greece, the United Kingdom, Czech Republic, Finland, Sweden, Denmark, or Korea. In these countries, grade retention rates are far below the OECD average of 11%. On the other hand, in countries such as Portugal, Spain, Belgium, Luxembourg, and Colombia, around 25% of the students reported having been retained at least once during their school career.

The difference in the retention rates reflects, thus, the different pedagogical strategies countries apply to manage students' heterogeneity in terms of achievement. According to Mons's (2007) typology, grade retention is not allowed or rarely applied in those countries that privilege the use of differentiation and individualised instruction over grade retention (individualised integration countries). Examples are Denmark, Iceland, Finland, and Norway, presenting relatively low retention rates (see Figure 2). Countries in a second group are those with 'à la carte integration' models (Mons, 2007), where grade retention is used as a last remedy when other measures, such as ability grouping or workload extension, were found insufficient (Dupriez et al., 2008). Thus, grade retention in these countries (e.g., United Kingdom, United States, Canada, New Zeeland, and Australia) are relatively low.

On the other hand, in 'separation countries' (Mons, 2007), grade retention is often applied together with other measures, such as separating students to different routes or tracks and early selection based on ability. Examples are Germany, Austria, Hungary, Switzerland, Luxembourg, Belgium, and Netherlands, whose rates are generally higher than the OECD average. Finally, Portugal is among the countries considered to have a 'uniform integration model' (Mons, 2007), along with France, Spain, Italy, and Colombia. These countries are recognised as using retention as the primary measure to deal with students' achievement, which is confirmed by their extremely high retention rates (Dupriez et al., 2008).

Figure 2

Percentage of 15-Year-old Students Retained at Least Once During Primary or Secondary Education, per OECD Country



The differences in the application of grade retention are also underpinned by conceptions and attitudes regarding students' development that, thus, lead to considering

grade retention as either a helpful or harmful intervention (Martin, 2011). Proponents of grade retention often rely on the nativist, or 'internal clock' view of child development (Martin, 2009, 2011), in which it is considered that students develop at their own pace. Thus, grade retention would grant students more time to mature, develop and catch up with the learning material without feeling frustrated and without having to work beyond their abilities (Goos, Schreirer et al., 2013; Hong & Yu, 2008; McCoy & Reynolds, 1999; Martin, 2011; Smith & Shepard, 1988; Tomchin & Impara, 1992; Wu et al., 2010).

Second, from the teachers' perspective, retaining students allow creating of more homogenous classrooms in terms of achievement, which would ease classroom instruction and would help teachers to meet students' individual needs (Dupriez et al., 2008; Hong & Raudenbush, 2005; Le Donné, 2014; Nunes et al., 2018; Shepard & Smith, 1989; Xia & Kirby, 2009). Third, teachers may often rely on their decision to retain a student on the potential benefits it would bring to their feelings of self-confidence and self-worth, as predicted by self-efficacy and the social comparison frameworks (Bandura, 1977; Festinger, 1954). Grade retention may constitute an opportunity for a fresh start. During the repeated year, students may finally experience success, feel more competent compared to their colleagues, gain positive feedback from teachers and stand out in the classroom in a positive way (Gleason et al., 2007). Altogether, these may stimulate students' self-concept, self-esteem, and motivation (Ehmke et al., 2017; Goos, Van Damme et al., 2013; Marsh, 2016; Marsh et al., 2017; Smith et al., 2022).

Finally, it is believed that the threat of grade retention would motivate students to work harder and to stay on track (Belot & Vandenberghe, 2014; Dupriez et al., 2008; Hong & Raudenbush, 2005; Range et al., 2012). In addition, it sends a message to students that lack of effort and poor performance has its consequences (Thompson & Cunningham, 2000; Valbuena et al., 2021).

Opponents of grade retention, however, argue that this practice does not bring benefits to students by considering the transitional (Jimerson, 2001), bioecological (Bronffenbrenner, 2006), sociocultural (Vygotsky, 1986), self-determination (Deci & Ryan, 2000), and stage-environment fit (Eccles & Midgley, 1989) theories of development and education. First, in most cases, grade retention mainly entails merely repeating the same learning material, impeding students from experiencing new challenges and academic growth. In addition, this mere repetition does not exactly tackle the students' difficulties that made them be retained in the first place. Altogether, this may undermine students' motivation (Goos, Van Damme et al., 2013; Ikeda & Garcia, 2014; Martin, 2011; Hong & Yu, 2007; Valbuena et al., 2021).

Second, students witness their former age-mates advancing to the next grade while they stay back in the same grade and are placed in a classroom with younger peers. Hence, this situation may trigger feelings of not fitting in, being out-of-place and not being accepted,

and disrupting students' social relationships. In addition, before retention, students may have received negative feedback from their teachers, given their poor performance or adjustment. At the same time, after retention, their teachers may also hold negative perceptions and lower expectations of them, treating them differently from their peers. Altogether, the stigma associated with being retained and the potential feelings of having failed could contribute to lower motivation, engagement, self-concept and self-esteem, and lower peer acceptance (Demanet & Van Houtte, 2016; Goos, Van Damme et al., 2013; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Pagani et al., 2001; Van Canegem et al., 2021; Wu et al., 2010; Xiang & Chiu, 2022).

Third, grade retention is perceived as a costly intervention (Brophy, 2006; Ehmke et al., 2010; European Commission, 2020; Flores et al., 2013; Higgins et al., 2022; Ikeda, 2011; Ikeda & Garcia, 2014; Justino et al., 2014; Valbuena et al., 2021), entailing both financial, opportunity and social costs for the students, their families, and the education system and society. For the education system, a student who is retained has the same economic effect as a new student entering the school system, doubling the financial cost and resources employed (Brophy, 2006; Flores et al., 213; Ikeda, 2011; Justino et al., 2014). In addition, it suggests that the resources provided were inefficient for students learning (Brophy, 2006). For students' families, it also requires extra financial support, at least for one year (Kretschmann et al., 2019). For students and society in broader terms, retention means delaying the entry into the labour market for at least one year (Ikeda, 2011) or into higher levels of education (European Commission, 2020), as well as potential risks of early school leaving (Ikeda, 2011; Jimerson, 1999; Jimerson et al., 2002).

Finally, it has been claimed that grade retention disproportionately affects students with different backgrounds and characteristics, especially the already academically disadvantaged students, questioning its fairness (European Commission, 2020; OECD, 2020). Thus, grade retention may exacerbate inequalities among students, and most retained students may not beneficiate from it (Dupriez et al., 2008; European Commission, 2020; Ikeda, 2011; OECD, 2020; Xiang & Chiu, 2022). With this in mind, several efforts have been made internationally to reduce grade retention rates.

Indeed, a decrease in retention rates was observed during the last decade. For example, Portuguese retention rates obtained from PISA in 2008 were around 35%, and the OECD average was 13% (Ikeda, 2011). This decrease suggests that countries are making significant efforts to reduce the use of retention (OECD, 2020). However, at the same time, the still considerable rates observed in many countries and the variation between countries under the same 'heterogeneity model' are signaling that other considerations, such as the characteristics of the student population, the school context, as well as the beliefs regarding retention (Agasisti & Cordero, 2017; Demanet & Van Houte, 2013; Eurydice, 2011; Goos,

Schreirer, et al., 2013; Young et al., 2019) may outweigh international recommendations for reducing retention.

1.2. Who is retained?

The most apparent reason to decide to retain students would be their poor academic achievement or not having reached the academic standards defined for a certain grade level (Agasisti & Cordero, 2017; Bonvin et al., 2008; OECD, 2020). However, studies consistently showed that other students' individual and contextual factors increase their likelihood of being retained. Identifying the determinants of grade retention is crucial to help signal students at greater risk for grade retention (Davoudzadeh et al., 2015). Such characteristics should also be considered when researching grade retention outcomes (Goos, Van Damme et al., 2013). These characteristics could be organized into four factors: students' demographics and family background, academic dispositions, behavioural dispositions, and school and system contexts (Agasisti & Cordero, 2017; Alexander et al., 2003).

In general, studies were fairly consistent in showing that grade retention is associated with students' demographics and family background. Specifically, the risk of grade retention increases for male students (Alexander et al., 2003; Choi et al., 2018; Davoudzadeh et al., 2015; Hong and Raudenbush, 2005; Hong and Yu, 2007; Huang, 2014; Jerrim et al., 2021; Jimerson et al., 1997; Klapproth & Schatlz, 2015; Pedraja-Chaparro et al., 2015), for those younger, and born at the end of the year (Choi et al., 2018; García-Perez et al., 2014; Huang, 2014; Jerrim et al., 2021; Pedraja-Chaparro et al., 2015), and for those from an ethnic minority or with an immigrant status (Alexander et al., 2003; Bali et al., 2005; Hong and Yu, 2007; Jimerson et al., 1997; Klapproth & Schaltz, 2015; Pedraja-Chaparro et al., 2015). Students' socioeconomic background and their family situation matter as well. Students with lower socioeconomic status are more likely to repeat a grade (Alexander et al., 2003; Bali et al., 2005; Choi et al., 2018: Davoudzadeh et al., 2015; Fine & Davis, 2003; Jerrim et al., 2021; Klapproth & Schaltz, 2015; Pedraja-Chaparro et al., 2015), as well as students with parents, particularly mothers with lower educational levels (Agasisti & Cordero, 2017; Alexander et al., 2003; Davoudzadeh et al., 2015; Gadeyne et al., 2008). The lower parental engagement was also associated with higher grade retention (Gadeyne et al., 2008; Hong & Raudenbush, 2005; Huang, 2014; Jimerson et al., 1997). In Portugal, similar profiles of retained students were found (Bastos & Ferrão, 2019; Conboy, 2011; Conboy et al., 2013; Ferrão, 2015; Nunes et al., 2018; Pereira & Reis, 2014).

Naturally, academic and cognitive variables also impact students' retention probability. Kindergarten attendance has been found to prevent students from being retained (Agasisti & Cordero, 2017; Davoudzadeh et al., 2015; Pedraja-Chaparro et al., 2015). Students are more likely to repeat a grade when they show lower levels of achievement (Bali et al., 2005; Bonvin

et al., 2008; Choi et al., 2018; Fine & Davis, 2003; Gadeyne et al., 2008; Huang, 2014; Jimerson et al., 1997; McCoy & Reynolds, 1999), when they are assessed as having lower cognitive abilities (Agasisti & Cordero, 2017; Bonvin et al., 2008; Huang, 2014; Jimerson et al., 1997; McCoy & Reynolds, 1999) and lower maturity (Bonvin et al., 2008; Huang, 2014; Pereira & Reis, 2014). Moreover, grade retention was associated with lower classroom adjustment and academic-related behaviours (Alexander et al., 2003; Gadeyne et al., 2008; Huang, 2014; Jimerson et al., 1997), lower self-concept and motivation (Ferrão, 2015; Kretschmann et al., 2019), and lower social and emotional skills, such as peer relationships or deviant behaviours (Demanet & Van Houtte, 2013; Gadeyne et al., 2008; Lubbers et al., 2006; Huang, 2014; Jimerson et al., 1997; Sidali et al., 2021).

The school context also matters in explaining grade retention practice, and one school factor, in particular, has consistently emerged from the literature. In schools with a higher disadvantaged composition, i.e., with a more significant proportion of disadvantaged students or sited in a disadvantaged area, students present a higher probability of being retained (Agasisti & Cordero, 2017; Davoudzadeh et al., 2015; Ferrão, 2015; OECD, 2020; Pires et al., 2021). In some contexts, such as the Portuguese and Spanish, attending a public school also increases the odds of being retained (Choi et al., 2018; Conboy, 2011; Conboy et al., 2013; Ferrão, 2015; Pires et al., 2021). This relates to the previous findings since private schools are most likely attended by socioeconomically advantaged students. Thus, disadvantaged students' likelihood of being retained is two-fold: they are more likely to be retained given their individual and family economic and social situation, and, in addition, because they are more likely to attend a school with a more disadvantaged composition. Caution is advised since grade retention may induce more inequalities between students (Conboy et al., 2013; Ferrão, 2015; OECD, 2020; Xiang & Chiu, 202).

Although in a limited number, the literature has also pointed to teachers' regarding the effectiveness of grade retention as a predictor of its use. Specifically, the more teachers believe that grade retention produces positive effects, the more they will retain their students (Bonvin et al., 2008; Santos et al., 2023). Furthermore, previous studies have shown that many teachers hold positive beliefs regarding grade retention effectiveness (Range et al., 2012; Santos et al., 2023; Tomchin & Impara, 1992). Finally, teachers' judgments of students, especially regarding their academic performance, competencies, maturity, or academic effort (Bonvin et al., 2008; Tomchin & Impara, 1992; Young et al., 2019), also play a crucial role in grade retention decision.

Taken together, these groups of determinants suggest that other variables are at play when teachers or schools decide to (or not) retain a student (Bonvin et al., 2008; OECD, 2020; Young et al., 2019). Thus, they must be considered not only when deciding in favour or against retaining a student but also when assessing the effects of retention.

1.3. Effects of grade retention

Given its controversial nature and its widespread application, the effects of grade retention have been extensively investigated and summarised (Allen et al., 2009; Bright, 2011; Holmes, 1989, Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009). Research on the effects of grade retention has a long tradition, as the first studies date back to the early 1900s (Jackson, 1975). Jackson (1975) conducted the first review on the effects of grade retention on students' academic achievement and personal and social adjustment. By reviewing 44 studies conducted between 1911 and 1973, Jackson (1975) concluded that the effects of grade retention did not outweigh the effects of grade promotion. Thus, he stated that "those educators who retain pupils in a grade do so without valid research evidence to indicate that such treatment will provide greater benefits to students with academic or adjustment difficulties than will promotion to the next grade (Jackson. 1975, p.627). In addition, Jackson (1975) questioned the quality of some research conducted to estimate grade retention effects and recommended using high-quality designs to compare retained students with their promoted counterparts.

Nearly one decade after the first review on grade retention effects, Holmes and Matthews (1984) conducted a meta-analysis of 44 studies published between 1929 and 1981 considering the effects of grade retention in primary and secondary education on students' achievement and socioemotional outcomes. Overall, they found an effect size (ES) of -0.37, suggesting detrimental effects of grade retention on students' achievement (ES = -0.44) and students' socioemotional outcomes (ES = -0.27). The review of Holmes (1989) updated the previous meta-analysis by reviewing 63 studies conducted between 1929 and 1989. In this review, the effects of grade retention were also generally adverse (ES = -0.19) on students' achievement (ES = -0.19) and socioemotional outcomes (ES = -0.09). Holmes (1989) called attention to the fact that the few (9) studies presenting positive effects "involved positive remediation plus retention" (p. 28).

Jimerson (2001) summarised the research findings of grade retention effects on both achievement and socioemotional variables, considering 20 studies investigating grade retention in primary and lower secondary education. The results obtained led Jimerson (2001) to conclude that grade retention is not an effective intervention for students' achievement (ES = -0.39) and socioemotional adjustment (ES = -0.22). In addition, the review from Jimerson et al. (2002), considering studies published between 1970 and 2000 on the relationship between grade retention and school dropout, found that "early grade retention is one of the most powerful predictors of later school withdrawal" (p. 452).

The effects of grade retention on students' academic achievement have been further summarised in Allen et al.'s (2009) and Bright's (2011) meta-analyses. Allen et al. (2009) challenged the quality of methods employed in previous reviews, especially the methodological

approaches of the reviewed studies. Thus, considering only studies that provided solid support for causal inferences, the results from 19 studies published between 1990 and 2007 yielded an overall null effect of grade retention (ES = 0.04). However, as Allen et al. (2009) mentioned, a nonsignificant finding does not favour grade retention. Instead, “a finding of “no significant difference” for retention on achievement calls into question the educational benefits of grade retention policies” (Allen et al., 2009, p. 492). In addition, the authors tested the moderating effect of the length of the follow-up and generally found more negative results as the years post-retention increased. Bright’s (2011) meta-analysis of studies conducted between 1990 and 2010 showed a more detrimental effect of retention on students’ achievement (ES = -0.50). However, the only methodological inclusion criterion was whether the study presented or not a comparison group.

At this point, the described reviews mainly focused on students’ achievement and, to a lesser extent, on students’ psychosocial adjustment and dropout. The review from Xia and Kirby (2009) expanded the scope by considering grade retention effects on other academic and non-academic variables, such as students’ postsecondary education enrollment and employment outcomes. In their literature review, 91 studies between 1980 and 2008 examined the effects of grade retention on students’ academic achievement, socioemotional outcomes and long-term educational (school dropout and postsecondary education enrollment) and employment outcomes. Overall, the authors found that the potential gains in achievement or socioemotional factors are very short-lived and tend to vanish over time (Xia & Kirby, 2009). Notably, this idea is emphasised by the long-term adverse effects of retention on students’ propensity to dropout of school and postsecondary education enrolment and on students’ employment outcomes in adulthood (Xia & Kirby, 2009). However, this narrative review pointed to the negligible number of studies considering the long-term academic and employment outcomes (Xia & Kirby, 2009).

More recently and in the course of this dissertation, Valbuena and colleagues (2021) conducted a systematic review aiming to provide evidence on grade retention in kindergarten, primary, and secondary education on students’ achievement, educational and employment outcomes. This review contained 42 studies using causal inference techniques published between 2001 and 2020. The findings of this review are in line with those from Xia and Kirby (2009), suggesting that benefits of grade retention, found mostly in short-term and under specific circumstances as such remediation interventions plus retention (Jackson, 1975), tend to vanish with academic and employment outcomes in outlook.

Summing up, the findings from these reviews generally questioned the effectiveness of grade retention (Allen et al., 2009; Bright, 2011; Holmes, 1989, Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009), given its most negative or null effects. Surprisingly, most of the studies reviewed so far were conducted in

the United States, except for a couple of papers reviewed by Xia & Kirby (2009) and Bright (2011) and, more recently, by Valbuena et al. (2021). In addition, none of the reviews evaluated grade retention's effects considering differences between countries with higher or lower retention rates.

Since grade retention is often a result of poor achievement, studying its effects on students' further academic development has been of concern for many researchers, educators and policymakers (Agasisti & Cordero, 2017; Bonvin et al., 2008; OECD, 2020). There are some reasons to expect that grade retention will improve students' achievement, as mentioned in section 1.1. On the one hand, grade retention is intended to provide students with the so-called 'gift of time' (Hong & Yu, 2008; Smith & Shepard, 1988; Wu et al., 2010), providing them with an additional year to master the learning material that they were not capable to do during the previous year. In addition, from a cognitive development perspective (Piaget & Inhelder, 1966), students only will be ready to master new learning material after consolidating the previous knowledge. On the other hand, grade retention allows creating of more homogeneous classrooms in terms of academic competencies, facilitating teacher instruction, and reducing the gap between students (Dupriez et al., 2008; Hong & Raudenbush, 2005; Le Donné, 2014; Nunes et al., 2018; Shepard & Smith, 1989; Xia & Kirby, 2009). At the same time, negative effects of grade retention may be found since it mainly consists of a mere repetition of all subjects, including those students that have succeeded. Thus, as predicted by sociocultural development theory (Vygotsky, 1986), such repetition deprives students of access to meaningful and stimulating intellectual challenges and learning, which are critical for cognitive development.

Considering the empirical findings considering the European context, some studies found positive short-term effects of grade retention (Goos, Van Damme, et al., 2013; Klapproth et al., 2016). Yet, in line with previous meta-analyses conducted in the United States, these initial positive effects diminish or even disappear in longer run (Alet et al., 2013; Goos, Van Damme et al., 2013; Ehmke et al., 2010, 2017; Klapproth et al., 2016; Vandecandelaere, Schmitt, et al., 2016). Similar findings were obtained in studies using Portuguese samples. Nunes et al. (2018) found no effects of repeating the 4th grade on students' 6th grade national exams. Pereira and Reis (2014) found negative effects of grade retention in primary education and positive effects of grade retention in secondary education on PISA reading and math scores, illustrating these short vs long-term differences. It is noticeable that teachers often whiteness only the temporary gains, and mainly because of these short-term effects on achievement, they tend to hold favourable beliefs regarding retention (Alet et al., 2013; Fine & Davis, 2003).

Although the effects of grade retention on students' academic achievement are relatively well established (Allen et al., 2009; Valbuena et al., 2021), the findings on the

relationship between grade retention and students' affective components of learning have been less consistent. One main reason for mixed findings is that the terms 'psychosocial outcomes' and 'socioemotional factors' are used interchangeably and cover a broad set of affective components of learning and other general social and emotional behaviours and attitudes. Jimerson (2001), for example, considered in his review that socioemotional variables were those that "included peer competence, self-esteem, locus of control, achievement expectations, school satisfaction, school engagement, behaviour problems, and other composite variables incorporating students' attitudes, behaviours, and social and emotional adjustment" (p. 424). On the other hand, Xia and Kirby (2009) in their review, considered socioemotional outcomes the following variables: "social adjustment, peer acceptance, perceived academic competence, self-concept, self-esteem/self-worth, emotional health, suicide, and student attitudes" (p. 22). These two definitions illustrate the complexity of drawing solid conclusions regarding grade retention's effects on these outcomes. In this work, we intended to contribute to this topic by considering specific affective components of learning often related to students' achievement and academic success: academic self-concept, self-esteem, motivational orientations, and school belonging.

Students' academic self-concept is defined as a cognitive self-evaluation of competence in specific academic-related domains (e.g., reading self-concept). This variable plays a significant and reciprocal role in students' school adjustment, achievement, and educational success (Guay et al., 2010; Huang, 2011; Marsh & Yeung, 1997; Shavelson et al., 1976). Conversely, self-esteem is considered an overall (negative or positive) affective judgment of one's self-worth (Harter, 1999).

Regarding students' motivational orientations, it is recognised that it is critical for students' academic behaviour and performance (Meece et al., 2006). Students' goal orientations have become one of the largest fields of research in students' motivation and can be characterised by students' reasons or purposes for engaging in certain achievement behaviours (Anderman & Wolters, 2006; Kaplan & Maher, 2007; Meece et al., 2006; Eccles & Wigfield, 2002; Skaalvik, 1997). Also interconnected with and often referred to as a social aspect of motivation (Goodenow & Grady, 1993; Martin et al., 2017), students' sense of school belonging or school attachment, school engagement, and identification with school (Finn, 1989; Goodenow & Grady, 1993; OECD, 2003; Voelkl, 1996; Slaten et al., 2016) is considered the feeling of being accepted, respected, valued, and supported by peers and the broader school community (Finn, 1989; Goodenow & Grady, 1993; OECD, 2003; Voelkl, 1996). These feelings are associated with feelings of valuing school and valuing school success. These two components of school belonging are closely linked with students' participation at school (Finn, 1989) and with other desirable educational outcomes, such as students' achievement,

motivation, academic engagement, and self-esteem (Goodenow & Grady, 1993; Ma, 2003; Martin et al., 2017; Sirin & Rogers-Sirin, 2004; Wentzel, 2002).

While grade retention is expected to affect students' achievement, it is also recognised that it may impact the abovementioned learning components. First, regarding academic self-concept and particularly to explain its positive effects, social comparison theory (Festinger, 1954), specifically the big-fish-little pond effect (Marsh, 1987; Marsh, Seaton, et al., 2008, Marsh, Trautwein, et al., 2008) has been often mentioned. In this regard, students form their self-perception of competence by comparing their school-related accomplishments with those of their classmates, and this frame of references acts as the base for their self-concept development (Marsh, 2016; Marsh et al., 2017; Marsh, Seaton, et al., 2008, Marsh, Trautwein, et al., 2008). After retention, students are placed in a classroom with younger and academically less experienced peers, which confers some advantages for repeaters. Thus, repeaters would adjust their competence beliefs by taking their new peers as the reference group (Marsh, 2016; Marsh et al., 2017).

Considering research findings on the effects of retention on students' self-concept, previous meta-analyses obtained an effect size of -0.19 (Holmes & Matthews, 1984), -0.13 (Holmes, 1989) or -0.04 (Jimerson, 2001). In addition, empirical studies within the European context also revealed mixed findings, revealing both positive (Ehmke et al., 2010; Lamote et al., 2016; Marsh et al., 2017) and adverse short-term effects (Kretschmann et al., 2019), while the long-term effects were mainly adverse (Klapproth et al., 2016; Van Canegem et al., 2021) or nonsignificant (Lamote et al., 2014; Marsh et al., 2017). In the Portuguese context, the results from Nascimento and Peixoto (2012) and Peixoto et al. (2016) showed negative effects of grade retention on students' self-concept in lower secondary education.

The negative findings may also be attributed to the fact that retained students might perceive that being retained constitutes a personal failure that makes them less competent and capable (Jimerson, 2001; Goos, Van Damme et al., 2013). Also, because of retention, students may receive a negative label such as 'slow learner' (Martin, 2011; Mathys et al., 2019), which contributes to the stigmatisation of these students within the school and classroom contexts (Demanet & Van Houtte, 2016; Wu et al., 2010). Altogether, these perceptions may undermine students' academic self-concept and self-esteem (Mathys et al., 2019; Nascimento & Peixoto, 2012; Peixoto et al., 2016).

To explain the effects of grade retention on students' motivation and engagement, the self-determination theory (Deci & Ryan, 2000), the stage-environment fit theory (Eccles & Midgley, 1989), and the self-worth theory (Covington, 1992) may provide valuable insights. Briefly, the self-determination theory (Deci & Ryan, 2000) posits that people have three psychological needs that must be fulfilled to promote their motivation: the need for competence, autonomy and relatedness (Deci & Ryan, 2000). In essence, the stage-

environment fit theory (Eccles & Midgley, 1989; Eccles et al., 1993) was elaborated to explain the decrease in students' motivation during adolescence. It is stated that students decline their academic motivation because they find a 'misfit' between their characteristics and psychological needs and the characteristics of their social environment (i.e., school). Finally, the self-worth theory (Covington, 1992, 2000) states that to protect their levels of self-acceptance, not well-succeeded students may engage in several avoidant behaviours to avoid failure, such as self-handicapping behaviours (e.g., procrastination) or disengagement (Castella et al., 2013; Covington, 2000).

Based on these frameworks, grade retention is expected to impact students' motivation and sense of belonging negatively (Demanet & Van Houtte, 2016; Goos, Van Damme et al., 2013; Ikeda & Garcia, 2014; Hong & Yu, 2007; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Pagani et al., 2001; Van Canegem et al., 2021; Wu et al., 2010; Xiang & Chiu, 2022). As already stated, grade retention is often only a matter of repetition of the same learning material at least one more time. Also, retained students might feel frustrated and as having failed their main mission: to succeed in school. After being retained, they are also detached from their peer group, with whom they most probably have formed significant relationships. Thus, altogether, these experiences may undermine students' needs for competence (e.g., self-concept), acceptance (e.g., belonging), as well students' adaptive goal orientations or success-oriented behaviours (Baumeister & Leary, 1995; Covington, 2000; Deci & Ryan, 2000; Eccles et al., 1993; Slaten et al., 2016). On the other hand, these experiences may also exacerbate feelings of isolation, being an outsider and being detached from school (Demanet & Van Houtte, 2016; Goos, Van Damme, et al., 2013). Moreover, grade retention is mainly involuntary, which may undermine students' sense of autonomy (Brophy, 2006).

Considering the effects of retention on students' motivation and belonging summarised in previous reviews, although they were not directly assessed, proximal terms are used, such as behavioural engagement and attitudes towards school. Regarding behavioural engagement as a proxy of students' motivation, the meta-analyses of Homes and Mathews (1984), Holmes (1989), and Jimerson (2001) showed mainly negative effects (ES =-0.31 to -0.11). Similarly, the effect size for attitudes towards school was, on average, -0.15 (Holmes, 1989; Holmes & Matthews, 1984). In what concerns empirical studies in Europe exploring these two variables, research is, to our knowledge, rather scarce. More specifically, the effects of grade retention on students' goal orientations are still understudied and, to our knowledge, only Nascimento and Peixoto (2012) and Peixoto et al. (2016, 2017) considered effects on this motivation framework, showing that retained students presented lower levels of task orientation (i.e., engagement in learning activities to develop task-related skills, Skaalvik, 1997) after being retained, and higher levels of avoidance orientation. In addition, considering students' sense

of belonging, Mathys et al. (2019) found negative short-term effects on secondary students', while Van Canegem et al. (2021) found negative effects, and Klapproth et al. (2016) found nonsignificant medium to long-term effects in the same age group.

Previous literature also claimed that to better understand grade retention effects, one should take into consideration the context under which it occurs (Demanet & Van Houtte, 2013, 2016; Hong & Raudenbush, 2005, 2006; Hong & Yu, 2008; Marsh, 2016; Van Canegem et al., 2021). Particularly, school retention composition through the proportion of retained students attending the same school has been found to affect all students' academic achievement (Bietenbeck, 2014; Gottfried, 2013; Hong & Raudenbush, 2005) and postsecondary enrollment intentions (Santos et al., 2022).

Considering the affective components of learning, Hong and Yu (2008) argued that "the social-emotional outcomes of a child are usually not independent of the social-emotional outcomes of his or her schoolmates" (p. 410). Thus, the spillover effects of retention on promoted students' psychosocial outcomes may also be expected (Bietenbeck, 2014; Gottfried, 2013). Indeed, the few existing studies showed that school retention composition affected students' same-grade friendships (Damenet & Van Houtte, 2016), school misconduct and academic behaviour (Bietenbeck, 2014; Demanet & Van Houtte, 2013), and school belonging (Van Canegem et al., 2021). These findings were justified by taking into consideration the fluctuation in retention rates across schools (Hong & Raudenbush, 2005) and the reference group theory (Richter, 1976) in that it is recognised that retained students attending a school with only a small number of peers in the same condition will more likely stand out in comparison with retained students in a school with a greater share of peers in the same situation (Van Canegem et al., 2021). Thus, grade retention is less stigmatising and has less detrimental effects in schools with a larger proportion of retained students (Demanet & Van Houtte, 2013). Further studies should consider this (moderating) variable in determining how grade retention experiences differ according to different school contexts. The impact of grade retention on repeaters' classmates will also be of interest since it is believed that it may undermine the classroom environment and non-retained students' performance (Lavy et al., 2012).

A final note must be given to the long-term effects of grade retention, particularly considering students' academic and professional career outcomes. Grade retention is often related to students' lower chances of high-school graduation and post-secondary enrolment (Belot & Vanderberghe, 2014; Cockx et al., 2019; Fine & Davis, 2003; Fraysier et al., 2020; Geng & Rockoff, 2017; Goos, Van Damme, et al., 2013; Huddleston, 2014; Jimerson,1999; Mendez et al., 2015; Santos et al., 2022), higher school-mobility, placement in special education and subsequent grade retention (Cockx et al., 2019; Geng & Rockoff, 2017; Goos, Van Damme, et al., 2013; Moser et al., 2012). In addition, grade retention was strongly

associated with students' school absenteeism and dropout (Allensworth, 2005; Eide & Showalter, 2001; Giano et al., 2022; Gubbels et al., 2019; Huddleston, 2014; Hughes et al., 2018; Jimerson, 1999; Jimerson et al., 2002; Ou & Reynolds, 2010). Regarding students' employment outcomes, the few existing studies suggest mainly nonsignificant or slightly negative effects (Baert & Picchio, 2021; Eide & Showalter, 2001; Jimerson, 1999; Ou & Reynolds, 2010) on students' earnings and professional career. Given these long-lasting detrimental effects of retention, Jimerson et al. (2002) stated that regarding retention, we may often win the battle (i.e., short-term benefit) but, in the end, lose the war (i.e., long-term detrimental effects). Studies within the Portuguese context considering these variables are very limited or inexistent in the case of professional career. The few studies investigating students' academic career showed that retained students demonstrated lower school progression (Nunes et al., 2018) and lower intentions to enrol in postsecondary education (Santos et al., 2022).

1.4. Grade retention in Portugal

The Portuguese education system comprises 12 years of compulsory education (Eurydice, 2022) and is organised in basic (grades 1 to 9) and secondary education (grades 10 to 12). Basic education is divided into three consecutive cycles: grades 1-4, corresponding to primary education; grades 5-6, corresponding to the second cycle; and grades 7-9, corresponding to lower secondary education. The Portuguese school system is categorised by Mons (2007) as a uniform system since the cycles of Basic education share a common core curriculum, whereas secondary education requires vocational choices (European Commission, 2020).

As mentioned previously and observed in Figure 2, grade retention is still widely applied in Portugal when students fail to reach academic standards established for a particular grade level. These high retention rates are related to the uniform structure of the Portuguese school system (Dupriez et al., 2008; Mons, 2007), in that, by sharing a common core curriculum, grade retention is the primary strategy used, acting both as a corrective and as a formative measure (Santos et al., 2023). As Dupriez et al. (2008) state, education systems with a uniform structure often lack other tools to deal with students' heterogeneity in terms of academic competencies, which also could be due to difficulties in finding alternative tools to deal with students' academic difficulties (European Commission, 2020).

Additionally, Portuguese high retention rates may reflect a 'culture of retention' (European Commission, 2020; Eurydice, 2011; Goos, Schreier, et al., 2013; Santos et al., 2023), meaning that teachers hold positive beliefs that grade retention will bring benefits to students. Recently, Santana (2019) and Santos et al. (2023) found in their studies that positive beliefs regarding grade retention are widespread across Portuguese school teachers. The

authors also found that these beliefs are closely linked to teachers' pedagogical beliefs, suggesting that teachers who hold transmissive models of learning tend to believe in the effectiveness of grade retention. Moreover, this culture of retention is stronger when positive beliefs are shared amongst teachers, and when teachers feel that their colleagues support their views, they tend to retain more students (Santos et al., 2023).

For over a decade, the Portuguese legislation has considered grade retention as an 'exceptional measure' that should only be used with a pedagogical purpose when grade promotion compromises the acquisition of new learnings (Decree-Law no. 55/2018, 2018, Portaria 223/A, 2018). The exception is for grade 1, where grade retention is not allowed (Decree-Law no. 55/2018, 2018, Portaria 223/A, 2018).

The main reasons to retain a student lie in achievement and attendance. The legislation in force defines that, for the end of each cycle (i.e., 4th, 6th, and 9th grade), students must be retained when they score insufficient, in primary education and below 3 (on a scale from 1 to 5) in second and third cycles, on core subjects or three school subjects (e.g., Portuguese, Math, and another compulsory subject; Portaria 223/A, 2018). In the remaining grade levels, grade retention decision falls on schools and on teachers who have (some) autonomy to define in which specific circumstances students are retained.

While the Portuguese retention rates stand out within the international context, a downward trend has been observed during the last decade in the number of students retained every year. Figures 3, 4 and 5 present grade retention and dropout rates in Basic education between the school year 2009/2010, representing the school year when students from Study 2 (chapter 4) enrolled in 1st grade, and the school year 2020/2021, the latest statistics available (Ministério da Educação/Direcção Geral de Educação, nd.).

Figure 3 clearly shows that the restrictions applied to 1st grade retention led to high retention in 2nd grade. However, grade retention is more common in the second and third cycle, as shown in Figures 4 and 5. These Figures also highlights the downward trajectory in grade retention rates starting in 2013 (CNE, 2020), presumably as a result of national policies implemented by the Portuguese Ministry of Education to reduce grade retention rates, such as 'Programa Mais Sucesso Escolar' (Barata et al., 2015). It should be mentioned that the lowest rates observed so far (the school year 2019/2020) correspond to the start of the COVID-19 pandemic and the related school closure and lockdown, which disrupted teaching and learning. Based on this, the National Association of School Psychologists ([NASP], 2021) advised against the use of retention as a measure to tackle learning losses during COVID-19. The same 'rule' was generally applied in Portugal, as seen by the lower grade retention rates. Nevertheless, after this disrupting period, the retention rates have risen by at least one percentage point (the school year 2020/2021).

Figure 3

Grade Retention and Dropout Rates Between 2010 and 2021 in the First Cycle

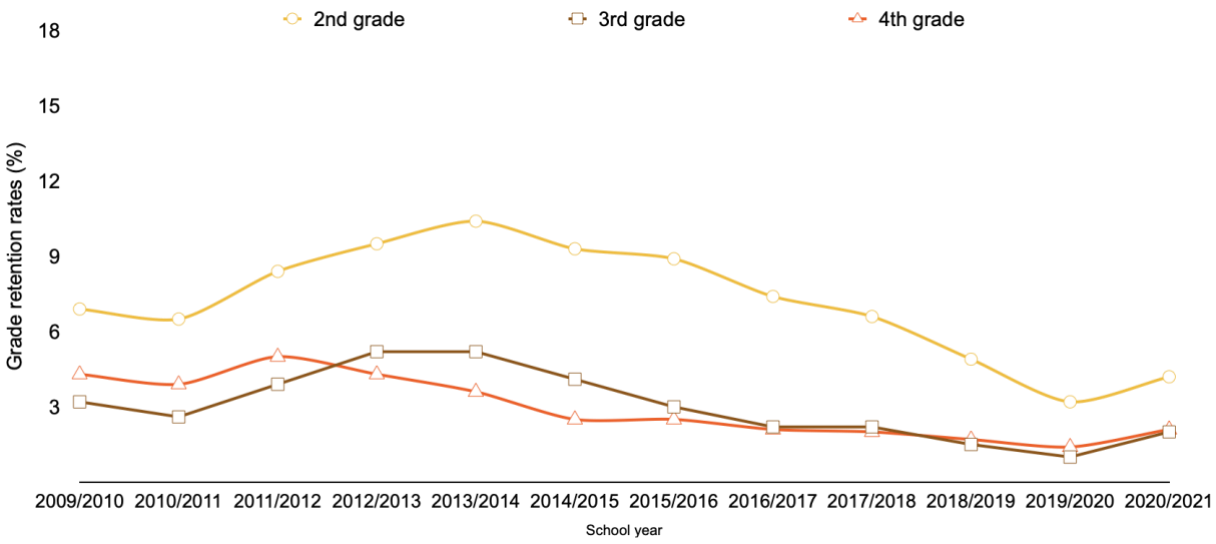
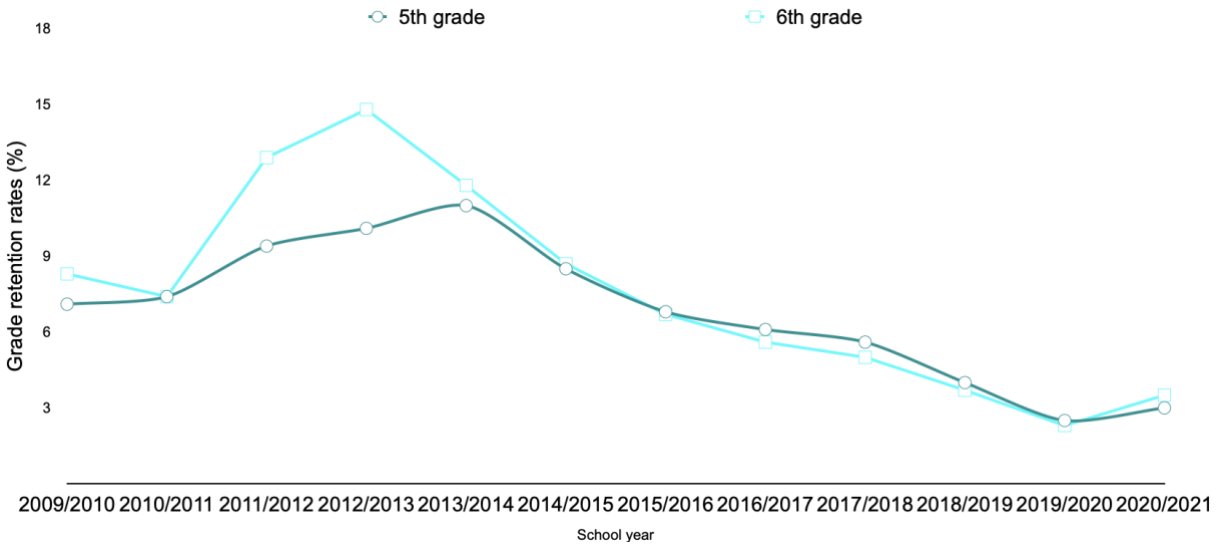


Figure 4

Grade Retention and Dropout Rates Between 2010 and 2021 in the Second Cycle

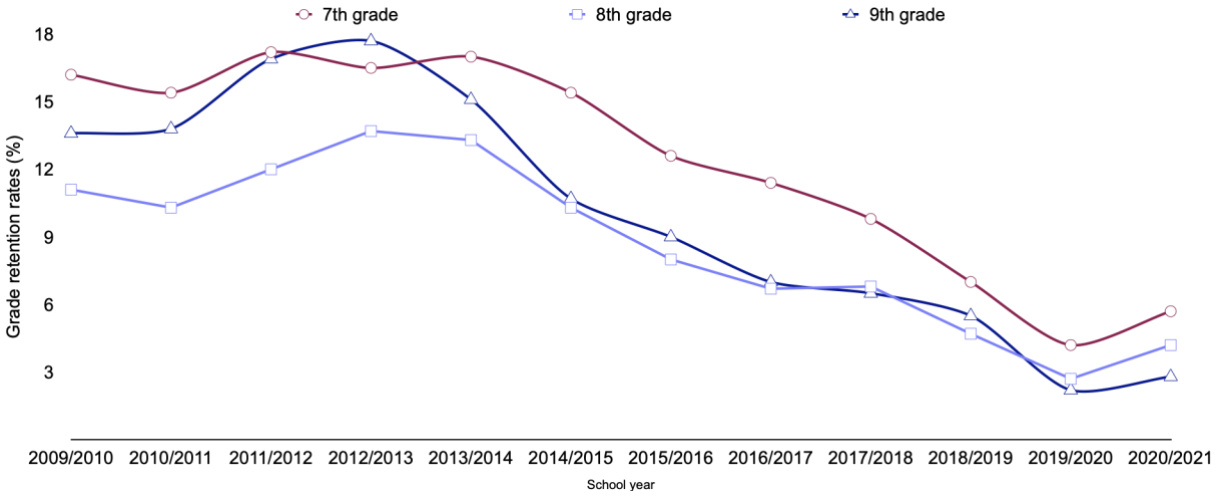


Notwithstanding the downward tendency of its application, grade retention is still often seen by teachers as a pedagogical and effective practice. Surprisingly, studies considering the effects of grade retention conducted in the Portuguese context still need to be made available. To the best of our knowledge, only a few studies considered the effects of grade retention on Portuguese students' achievement (Conboy et al., 2013; Nunes et al., 2018; Pereira & Reis, 2014; Rosário et al., 2013), affective components of learning (Nascimento & Peixoto, 2012;

Peixoto et al., 2016, 2017), and students' school career (Nunes et al., 2018; Santos et al., 2022).

Figure 5

Grade Retention and Dropout Rates Between 2010 and 2021 In the Third Cycle



2. THE PRESENT DISSERTATION

2.1. Research aims

Grade retention is a widely applied, as well as debated, educational measure. This debate is particularly fueled by mixed findings regarding grade retention effects on students' achievement, psychosocial factors, academic career and professional career (Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009). Grade retention deals with one of the most precious things one has: time. Thus, it would be critical to understand whether, by retaining students in grade, we would be providing them more time or taking that time from them (Ehmke et al., 2010; Ikeda, 2011; Valbuena et al., 2021). In addition to these timely costs, grade retention also has economic, social and opportunity costs (Brophy, 2006; Ehmke et al., 2010; European Commission, 2020; Flores et al., 2013; Higgins et al., 2022; Ikeda, 2011; Ikeda & Garcia, 2014; Justino et al., 2014; Valbuena et al., 2021). Grade retention is often negatively associated with equity in Education (European Commission, 2020; Lavy et al., 2012; OECD, 2020; Xiang & Chiu, 2022), given that it may exacerbate student inequalities.

Notwithstanding these costs, grade retention is still applied, and researchers still seek the answer to its effectiveness. Several authors have advised the research community to use sound methodologies to provide causal inferences of grade retention effects (Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2001; Xia & Kirby, 2009). Yet, several studies do not mirror this recommendation, which may contribute to the less consensus around retention. Also, the answer to the question of how grade retention affects students is rather complex because it is affected by and affects different aspects of students' lives (Agasisti & Cordero, 2017; Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2001; Xia & Kirby, 2009). Notably, most review studies that sought to provide clarity regarding grade retention effectiveness have directed little attention to other variables in addition to students' achievement, such as psychosocial, academic, and professional career outcomes (Allen et al., 2009; Bright, 2011; Holmes, 1989; Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Xia & Kirby, 2009).

Naturally, the effects of grade retention on students' achievement and their academic career should not be understood without the affective components of learning (Jimerson, 2001), given the bidirectional relationship between these components, such as motivation, engagement, self-concept, and their achievement and school success (Goodenow & Grady, 1993; Guay et al., 2010; Huang, 2011; Ma, 2003; Marsh, 2016; Marsh & Yeung, 1997; Martin et al., 2017; Shavelson et al., 1976; Sirin & Rogers-Sirin, 2004). Even when debating about the benefits or detrimental effects of grade retention, the main reasons raised are related to affective factors, such as students' self-esteem, motivation, self-confidence, etc. (Dupriez et al., 2008; Ehmke et al., 2017; Goos, Schreirer et al., 2013; Hong & Raudenbush, 2005; Hong & Yu, 2008; McCoy & Reynolds, 1999; Marsh, 2016; Marsh et al., 2017; Martin, 2011; Nunes

et al., 2018; Shepard & Smith, 1989; Smith & Shepard, 1988; Tomchin & Impara, 1992; Wu et al., 2010; Xia & Kirby, 2009; Young et al., 2019).

Given this, the first general aim of this dissertation was to investigate the effectiveness of grade retention by mapping and synthesising the research findings from studies conducted during the last two decades on the effects of grade retention on students' achievement, psychosocial variables, academic career, and professional career. Moreover, given the already mentioned importance of the affective components of learning on students' academic success, and the less consensual findings regarding the effects of grade retention on this group of variables, the second general aim of this dissertation was to investigate the effects of grade retention on students' affective components of learning. Considering the second general aim of this dissertation of investigating the affective components of learning, two specific aims were established:

Aim 2a: To investigate the effects of grade retention on (the development of) Portuguese secondary students' self-concept, self-esteem goal orientations, school belonging and school valuing;

Aim 2b: To investigate the moderating effects of school retention composition on the relationship between grade retention and those affective components.

The rationale behind Aim 2a was twofold: first, as already documented, the studies conducted within the Portuguese context are very limited, and only a few considered the affective components of learning (Nascimento & Peixoto, 2012; Peixoto et al., 2016, 2017). As was also already demonstrated, grade retention is a common practice in Portugal (see Figure 2). While its rates have been declining every year (see Figures 3, 4, and 5), teachers' beliefs regarding the effectiveness of this practice are most favourable (Santana, 2019; Santos et al., 2023). This scenario highlights the need for clarity regarding grade retention's effects in this specific context. Second, many studies have investigated grade retention effects in primary education (e.g., Goos, Van Damme, et al., 2013; Gleason et al., 2007; Hwang & Cappella, 2018; Moser et al., 2012; Vandecandelaere, Schmitt, et al., 2016; Wu et al., 2010). However, some researchers claimed that grade retention might be more harmful in later grades (Demanet & Van Houtte, 2016; García-Perez et al., 2014; Giano et al., 2022; Jimerson & Ferguson, 2007; Mathys et al., 2019; Wu et al., 2010) when the stigma of being retained may be more pronounced and when students may be more aware regarding its consequences (Anderson et al., 2005; Demanet & Van Houtte, 2013, 2016; Wu et al., 2010).

Aim 2b was grounded in the importance of the school context in understanding grade retention effects (Demanet & Van Houtte, 2013, 2016; Hong & Yu, 2008; Marsh, 2016; Smith & Shepard, 1988; Van Canegem et al., 2021). Following previous research on this matter (Bietenbeck, 2014; Demanet & Van Houtte, 2013, 2016; Gottfried, 2013; Hong & Raudenbush, 2005; Van Canegem et al., 2021), we considered school retention composition, meaning the

proportion of retained students attended to the same school. The research on this moderator is still scarce but has consistently shown that school retention composition affects students' psychosocial outcomes (Demanet & Van Houtte, 2013, 2016; Van Canegem et al., 2021). This limited number of studies suggests that more investigation is needed to unveil the impact of school retention composition on students' outcomes.

The two general aims of this dissertation and the two specific aims were addressed through three empirical studies, which will be briefly described in the next section. Finally, this dissertation aimed to inform researchers, teachers, the broader school community, and policymakers about the effectiveness of grade retention and to shed light on the next steps for providing support to struggling students.

2.2. Overview of the empirical studies

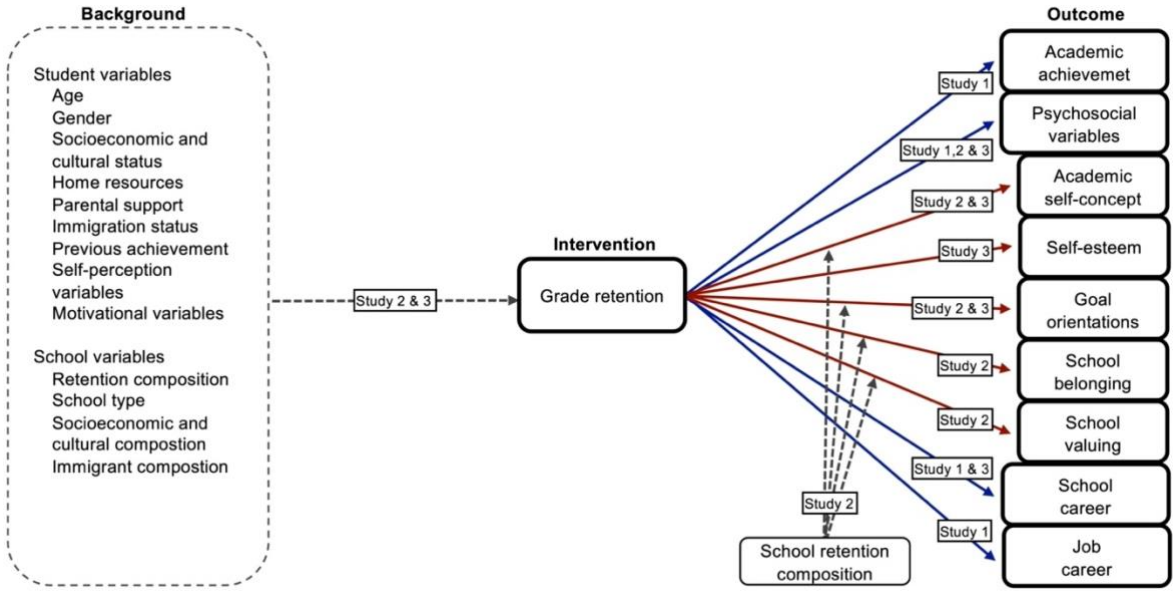
This dissertation presents three empirical studies that sought to bridge gaps identified in the literature review, considering the body of studies focused on studying grade retention effectiveness. Before moving to the empirical section, we provide a brief overview of the research aims of each study. Figure 6 presents the main variables addressed in the empirical studies.

In chapter 3, we present Study 1: a meta-analysis that investigated the effects of grade retention on several students' academic and non-academic outcomes, namely, academic achievement, psychosocial variables, school career, and professional career (see Figure 6). Additionally, we aimed to investigate whether the effects of grade retention differ by the context where it is applied (i.e., country), the retention application (i.e., whether additional interventions are provided), the outcome domain (i.e., the outcomes listed above), the outcome timing (e.g., short vs long-term results), the comparison approach (i.e., age vs grade comparison), and the research method employed (i.e., the quasi-experimental method used).

Studies 2 and 3 are presented in Chapters 4 and 5, respectively, and share a common focus, as seen in Figure 6: affective components of learning. Study 2 presented a retrospective cross-sectional study using the latest PISA data and aimed to investigate the effects of grade retention on secondary students' academic self-concept, goal orientations, school belonging and school utility value. Comparing students with a history of grade retention with their promoted peers of same age aimed to answer the following question: *how retained students would have scored on these variables had they been promoted instead of being retained?* Moreover, in this study, we aimed to investigate whether school retention composition (i.e., the proportion of retained students attending the same school) moderated the relationship between grade retention and the affective components of learning.

Figure 6

Overview of the Variables Under Study



Finally, in Study 3, we conducted a longitudinal study to understand the effects of grade retention in lower secondary education on students' trajectories of academic self-concept, self-esteem, and goal orientations. Additionally, in exploratory terms, we investigated the long-term effects of these variables on students' school career. Since retained students were compared with their age peers, we aimed to answer the same counterfactual question posited in Study 2.

2.3. Common methodological aspects

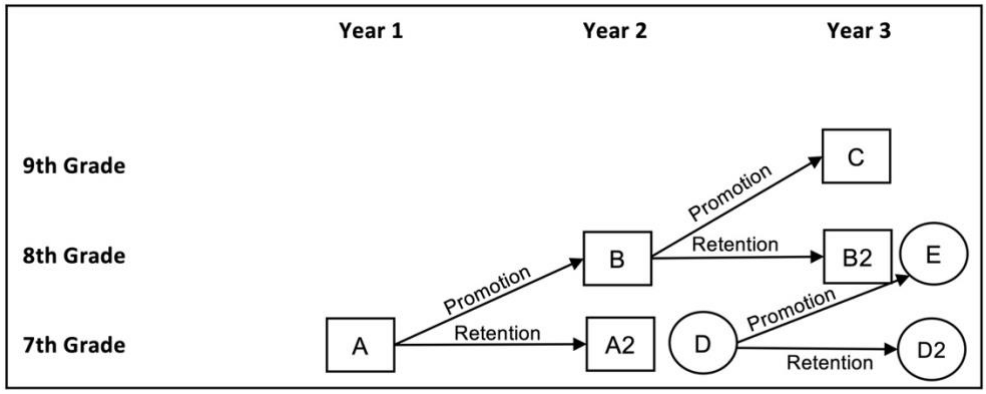
In our empirical studies, we made a great effort to follow the recommendations of recent studies investigating grade retention effects (e.g., Allen et al., 2009; Valbuena et al., 2021; Xia & Kirby, 2009). Thus, we made clear which type of comparison we used, on the one hand, and which design we applied. In the following paragraphs, we elaborate upon these methodological aspects.

2.3.1. Type of comparison

First, regarding the type of comparison, the literature identifies two possible comparisons between retained and promoted students: same-grade and same-age comparison (Hong & Raundenbush, 2005; Xia & Kirby, 2009). Figure 7 illustrates these two types of comparisons considering as an example of students' lower secondary school career that is addressed in Study 3.

Figure 7

Illustration of Comparison Approach Adapted From Xia & Kirby (2009)



In same-grade comparison, the outcome of students' repeated year is compared with their peers of the same grade (either their new, younger classmates or their former classmates a school year earlier; Xia & Kirby, 2009). In Figure 7, this means that the outcome of students that were retained either in 7th or 8th grade (A2 and B2 respectively) is compared with either the outcome of their former classmates when where they were in grade 7 (A) or grade 8 (B), i.e., when they were in that grade level for the first time. Another option is to compare retained students with their new younger classmates (D and E, respectively). In this type of comparison, studies sought to answer how repeaters score at the cost of an extra year compared to their younger same-grade peers with whom they will eventually share the rest of their school career (Hong & Raudenbush, 2005).

In same-age comparisons, repeaters are compared with their peers of the same age who advanced to the next grade level. Figure 7 illustrates this by the promotion paths (B for 7th grade and C for 8th grade). This comparison type aims to show how repeaters would fare had they been promoted instead of retained (Allen et al., 2009; Hong & Raudenbush, 2005; Hong & Yu, 2008).

Both comparisons are valid and provide valuable insights regarding grade retention, however, one should be aware that they produce different answers. Several researchers have stated that same-grade comparisons benefit retained students, particularly in the short term, as they are compared with younger and less experienced classmates (Allen et al., 2009; Hong & Raudenbush, 2005). On the other hand, same-age comparisons tend to find more favourable results for promoted students since they are continuously exposed to new challenges and are not detached from their class group (Allen et al., 2009; Hong & Raudenbush, 2005). In addition, as mentioned by Hong & Raudenbush (2005), studies using age comparisons tend to produce more consistent results, suggesting that promoted students usually present better outcomes than their same-age retained students.

Considering the approach used in the present work, Study 1 explored the moderating effects of grade vs age comparison approaches on the effects of grade retention. In both Study 2 and 3, we used a same-age comparison since we were interested in investigating the counterfactual question of how students would have developed if promoted instead of retained.

2.3.2. Methodological (statistical) approaches

Another methodological aspect often debated in grade retention literature (e.g., Allen et al., 2009; Valbuena et al., 2021; Xia & Kirby, 2009) is the necessity of finding a suitable comparison group of non-retained students. Naturally, comparing retained students with the *whole* group of non-retained students would be unfair to the former and would not produce valid and reliable conclusions regarding the effectiveness of grade retention (Alexander et al., 2003; Allen et al., 2009). Indeed, in any classroom, there is a substantial variation in students' achievement, competencies and affective experiences. Thus, comparing students with academic difficulties with those with the highest achievement levels would only create more disparities. Second, as already discussed, several characteristics could increase the risk of retention – grade retention predictors- and, again, it would not be appropriate to compare students at greater risk with those with little risk of retention that would naturally differ in these characteristics (Alexander et al., 2003; Hong & Raudenbush, 2005).

Causal inferences regarding a particular treatment or domain are, naturally, appropriately investigated using experimental research by randomly assigning students to treatment and comparison groups while controlling for several critical variables. However, most studies investigating grade retention effects are observational in nature. Randomly assigning students to a group of repeaters and non-repeaters would also be impossible for ethical reasons. Thus, to overcome these methodological issues, researchers have been relying on quasi-experimental methods (e.g., propensity score, differences-in-differences, regression discontinuity, and instrumental variables; for an overview, see Murnane & Willet, 2011) to estimate grade retention effects by comparing repeaters with promoted students with some degree of risk of being retained and presenting similarities in several pre-retention characteristics.

Mixed findings in grade retention research are often attributed to the methodological approaches (e.g., quasi-experimental vs regression or other methods; Allen et al., 2009; Valbuena et al., 2021; Xia & Kirby, 2009). Moreover, Allen et al. (2009) noted in their meta-analysis that less rigorous studies (i.e., using regression or simple comparison methods) reported more favourable effects of grade retention than quasi-experimental and more rigorous ones. Using quasi-experimental methods over other methods, such as regression methods with selected covariates, is preferable as it is more effective in dealing with selection bias and accounting for pre-treatment differences between retained and promoted students (Allen et al.,

2009; Murnane & Willlet, 2011). In addition, as Rubin (2001) mentioned, the study's design must be separated from the outcome analysis, supporting the application of quasi-experimental methods over regression analysis that controls for covariates.

In this dissertation, we followed the recommendations of previous research. In Study 1, we summarised the effects of grade retention of high-quality studies by clearly defining the inclusion of studies employing quasi-experimental methods as criteria. Additionally, we investigated the moderating effects of these different methods on the relationship between grade retention and outcomes.

In Studies 2 and 3, among the existing quasi-experimental methods, we used propensity score methods (Austin, 2011; Rosenbaum & Rubin, 1983; Stuart, 2010), an increasingly popular approach among studies estimating grade retention effects on psychosocial outcomes (e.g., Ehmke et al., 2010; Ehmke et al., 2017; Goos, Van Damme, et al., 2013; Hong & Yu, 2008; Hwang & Cappella, 2018; Im et al., 2013; Klapproth et al., 2016; Kretschmann et al., 2019; Lamote et al., 2014; Mathys et al., 2019; Vandecandelaere, Schmitt, et al., 2016; Wu et al., 2010). In the words of Rosenbaum and Rubin (1983), propensity score refers to the conditional probability to be retained given a set of observed covariates. Thus, before conducting the outcome analysis, we first resorted to different propensity score methods (Austin & Stuart, 2015; Stuart & Green, 2008; Vandecandelaere, Vansteelandt, et al., 2016) to assure comparability between retained and promoted students to reduce bias between retained and promoted students in background pre-treatment variables (see the left side of Figure X for an overview of background variables).

References

Agasisti, T., & Cordero, J. M. (2017). The determinants of repetition rates in Europe: Early skills or subsequent parents' help? Journal of Policy Modeling, *39*, 129-146. https://doi.org/10.1016/j.jpolmod.2016.07.002

Alet, É., Bonnal, L., & Favard, P. (2013). Repetition: Medicine for a short-run remission. *Annals of Economics and Statistics, 111*, 227-250. https://doi.org/10.2307/23646332

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary school grades*. Cambridge University Press.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A Meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis, 31*, 480–499. https://doi.org/10.3102/0162373709352239

Allensworth, E. M. (2005). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Educational Evaluation and Policy Analysis, 27*(4), 341-364. https://doi.org/10.3102/01623737027004341

Anderman, E., & Wolters, C. (2006). Goals, values and affect: Influences on students motivation. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 369-389). Lawrence Erlbaum Associates.

Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005) Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology, 21*, 1–20. https://doi.org/10.1300/J370v21n01_01

Austin, P. C. (2011). An introduction to propensity score methods for reducing confounding effects in observational studies. *Multivariate Behavioral Research, 46*, 399–424. https://doi.org/10.1080/0027 3171.2011.568786

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies*. Statistics in Medicine, 34*, 3661–3679. https://doi.org/10.1002/sim.6607

Baert, S., & Picchio, M. (2021). A signal of (train) ability? Grade repetition and hiring chances. *Journal of Economic Behavior & Organization, 188*, 867-878. https://doi.org/10.1016/j.jebo.2021.06.010

Bali, V. A., Anagnostopoulos, D., & Roberts, R. (2005). Toward a political explanation of grade retention. *Educational Evaluation and Policy Analysis*, *27*, 133–155. https://doi.org/10.3102/016237370270021

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Barata, M. C., Calheiros, M. M., Patrício, J. N., Graça, J., & Lima, M. L. (2015). Evaluating the impact of national educational policy to reduce retention and increase achievement in compulsory education. *The Elementary School Journal, 116*, 149–171. https://doi.org/10.1086/682748

Bastos, A., & Ferrão, M. E. (2019). Analysis of grade repetition through multilevel models: A study from Portugal. *Cadernos de Pesquisa, 49*, 270–288. https://doi.org/10.1590/198053146131

Baumeister, R., & Leary, M. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 17*, 497-529. http://doi.org/10.1037/0033-2909.117.3.497

Belot, M., & Vandenberghe, V. (2014). Evaluating the "threat" effects of grade repetition: Exploiting the 2001 reform by the French-speaking community of Belgium. *Education Economics, 22*, 73-89. https://doi.org/10.1080/09645292.2011.607266

Bietenbeck, J. (2014). *Learning from adversity? Short- and long-term spillover effects from grade retention in kindergarten*. https://www.researchgate.net/publication/268447733_Learning_from_Adversity_Short-and_Long-Term_Spillover_Effects_from_Grade_Retention_in_Kindergarten

Bonvin, P., Bless, G., & Schuepbach, M. (2008). Grade retention: Decision-making and effects on learning as well as social and emotional development. *School Effectiveness and School Improvement, 19*, 1-19. https://doi.org/10.1080/09243450701856499

Bright, A. D. (2011). *A meta-analysis of the effects of grade retention of K-6 students on student achievement: 1990-2010*. [Doctoral dissertation, University of Alabama]. Institutional Repository of the University of Alabama. http://acumen.lib.ua.edu/content/u0015/0000001/0000747/u0015_0000001_0000747

Bronfenbrenner U., & Morris P. A. (2006). The bioecological model of human development. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Theoretical model of human development* (pp. 793-828). John Wiley.

Brophy, J. (2006). *Grade repetition.* International Academy of Education. International Institute for Educational Planning.

Castella, K., Byrne, D., & Covington, M. (2013). Unmotivated or motivated to fail? A cross-cultural study of achievement motivation, fear of failure, and student disengagement. *Journal of Educational Psychology, 105*, 861–880. https://doi.org/10.1037/a0032464

Choi, A., Gil, M., Mediavilla, M., & Valbuena, J. (2018). Predictors and effects of grade repetition in Spain. *Revista de Economia Mundial, 48,* 21-42. https://doi.org/10.33776/rem.v0i48.3882

Cockx, B., Picchio, M., & Baert, S. (2019). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics, 34*, 403-424. https://doi.org/10.1002/jae.2670

Conboy, J. (2011). Retention and science performance in Portugal as evidenced by PISA. *Procedia Social and Behavioral Sciences, 12,* 311-321. https://doi.org/10.1016/j.sbspro.2011.02.040

Conboy, J., Moreira, I., Santos, I., & Fonseca, J. (2013). Práticas e consequências da retenção escolar: Alguns dados do PISA. In L. Veloso & P. Abrantes (Orgs.), *Sucesso escolar: Da compreensão do fenómeno às estratégias para o alcançar* (pp. 9-28). Mundos Sociais.

Conselho Nacional de Educação (2020). *Estado da educação 2019* [State of Education 2019]. Conselho Nacional de Educação. https://www.cnedu.pt/pt/publicacoes/estado-da-educacao/1615-estado-da-educacao-2020

Covington, M. V. (1992). Making the grade: A self-worth perspective on motivation and school reform. Cambridge University Press. https://doi.org/10.1017/CBO9781139173582.

Covington, M. V. (2000). Goal theory, motivation and school achievement: An integrative review. *Annual Review of Psychology, 51*, 171-200. https://doi.org/10.1146/annurev.psych.51.1.171

Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11,* 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Decree-Law No. 55/2018 (2018). Presidência do Conselho de Ministros. Diário da República – Série I – No. 129 de 6 de Julho, pp. 2928-2943. https://data.dre.pt/eli/dec-lei/55/2018/07/06/p/dre/pt/

Davoudzadeh, P., McTernan, M. L., and Grimm, K. J. (2015). Early school readiness predictors of grade retention from kindergarten through eighth grade: a multilevel discrete-time survival analysis approach. Early Child. Res. Q. 32, 183–192. https://doi.org/10.1016/j.ecresq.2015.04.005

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement, 24*, 417-434. https://doi.org/10.1080/09243453.2012.727834

Demanet, J., & Van Houtte, M. (2016). Are flunkers social outcasts? A multilevel study of grade retention effects on same-grade friendships. American Educational Research Journal, 53, 745– 780. https://doi.org/10.3102/0002831216646867

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils’ heterogeneity? *Comparative Education Review, 52*, 245–273. https://doi.org/10.1086/528764

Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate

classrooms for early adolescents. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 139-186). Academic Press.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *The American Psychologist, 48*, 90–101. https://doi.org/10.1037//0003-066x.48.2.90

Eccles, J., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review Psychology, 53*, 109- 132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Ehmke, T., Drechsel, B., & Carstensen, C. H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. *Studies in Educational Evaluation, 36*, 27–35. https://doi.org/10.1016/j.stueduc.2010.10.003

Ehmke, T., Salzer, C., Pietsch, M., Drechsel, B., & Muller, K. (2017). Competence development in the school year after PISA 2012: Effects of grade retention. *Zeitschrift fur Erziehungswissenschaft, 20*, 99–124. https://doi.org/10.1007/s11618-017-0752-4

Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. Economics of Education Review, 20(6), 563-576. https://doi.org/10.1016/S0272-7757(00)00041-8

European Commission. (2020*). Equity in school education in Europe. Structures, policies and student performance*. Eurydice report. Publications Office of the European Union. https://doi.org/10.2797/286306

Eurydice. (2011). Grade retention during compulsory education in Europe: Regulations and statistics (9789292011406). https://op.europa.eu/en/publication-detail/-/publication/10f14860-12a3-4f9e-b10a-32bee875420d

Eurydice. (2022). *Key features of the education system: Portugal*. https://eurydice.eacea.ec.europa.eu/national-education-systems/portugal/portugal

Ferrão, M. E. (2015). Tópicos sobre retenção escolar em Portugal através do PISA: Qualidade e equidade. *Arquivos Analíticos de Políticas Educativas, 23*, 1-19. http://www.redalyc.org/articulo.oa?id=275041389105

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140. http://dx.doi.org/10.1177/001872675400700202

Fine, J. G., & Davis, J. M. (2003). Grade retention and enrollment in post-secondary education. *Journal of School Psychology, 41*, 401-411. http://dx.doi.org/10.1016/j.jsp.2003.07.001

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117–142. https://doi.org/10.3102/00346543059002117

Flores, I., Mendes, R., & Velosa, P. (2013). O que se passa que os alunos não passam. In

Conselho Nacional de Educação (Ed.), *Estado da Educação 2013* (pp. 374-391). Conselho Nacional de Educação. https://www.cnedu.pt/content/edicoes/estado_da_educacao/Estado-da-Educacao-2013-online-v4.pdf

Fraysier, K., Reschly, A., & Appleton, J. (2020). Predicting postsecondary enrolment with secondary student engagement data. *Journal of Psychoeducational Assessment, 38*, 882–899. https://doi.org/10.1177/0734282920903168

Gadeyne, E., Onghena, P., & Ghesquière, P. (2008). child and family characteristics associated with nonpromotion in preprimary education. *Exceptional Children, 74*, 453–469. https://doi.org/10.1177/001440290807400403

García-Pérez, J. I., Hidalgo-Hidalgo, M., & Robles-Zurita, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied economics, 46*, 1373-1392. http://dx.doi.org/10.1080/00036846.2013.872761

Geng, T., & Rockoff, J. (2017). Does repeating a grade make students (and parents) happier? Regression discontinuity evidence from New York City. https://www.udesa.edu.ar/sites/default/files/rockoff.pdf

Giano, Z., Williams, A. L., & Becnel, J. N. (2022). Grade retention and school dropout: Comparing specific grade levels across childhood and early adolescence. *Journal of Early Adolescence, 42*, 33–57. https://doi.org/10.1177/02724316211010332

Gleason, K. A., Kwok, O. M., & Hughes, J. N. (2007). The short-term effect of grade retention on peer relations and academic performance of at-risk first graders. *The Elementary School Journal, 107*, 327–340. https://doi.org/10.1086/516667

Goodenow, C., & Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *The Journal of Experimental Education, 62*, 60–71. https://doi.org/10.1080/00220973.1993.9943831

Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review, 57*, 54-84. https://doi.org/10.1086/667655

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*, 323–347. https://doi.org/10.1016/j.jsp.2013.03.002

Gottfried, M. A. (2013). The spillover effects of grade-retained classmates: Evidence from urban elementary schools. *American Journal of Education, 119*, 405–444. https://doi.org/10.1086/669851

Guay, F., Ratelle, C. F., Roy, A., & Litalien, D. (2010). Academic self-concept, autonomous

academic motivation, and academic achievement: Mediating and additive effects. *Learning and Individual Differences, 20*, 644-653. https://doi.org/10.1016/j.lindif.2010.08.001

Gubbels, J., van der Put, C. E., & Assink, M. (2019). Risk factors for school absenteeism and dropout: A meta-analytic review. *Journal of Youth and Adolescence 48*, 1637–1667. https://doi.org/10.1007/s10964-019-01072-5

Harter, S. (1999). *The construction of the self: A developmental perspective.* Guilford Press.

Higgins, S., Katsipataki, M., Villanueva Aguilera, A. B., Dobson, E., Gascoine, L., Rajab, T., Kalambouka, A., Reardon, J., Stafford, J., & Uwimpuhwe, G. (2022). The Teaching and Learning Toolkit: Communicating research evidence to inform decision-making for policy and practice in education. *Review of Education, 10*, e3327. https://doi.org/10.1002/rev3.3327

Holmes, C. T. (1989). Grade-level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp.16-33). The Falmer Press.

Holmes, C. T., & Matthews, K. M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*, 225-236. https://doi.org/10.3102/00346543054002225

Hong, G. L., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*, 205–224. https://doi.org/10.3102/01623737027003205

Hong, G. L., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*, 901-910. doi:10.1198/016214506000000447

Hong, G. L., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis, 29*(4), 239-261. https://doi.org/10.3102/0162373707309073

Hong, G. L., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44,* 407–421. https://doi.org/10.1037/0012-1649.44.2.407

Huang, C. (2011). Self-concept and academic achievement: A meta-analysis of longitudinal Relations. *Journal of School Psychology, 49*, 505-528. https://doi.org/10.1016/j.jsp.2011.07.001

Huang, F. L. (2014). Further understanding factors associated with grade retention: Birthday effects and socioemotional skills. *Journal of Applied Developmental Psychology, 35*, 79-93. https://doi.org/10.1016/j.appdev.2013.12.004

Huddleston, A. P. (2014). Achievement at whose expense? A literature review of test-based grade retention policies in U.S. school. *Education Policy Analysis Archives, 22*, 1-31. https://doi.org/10.14507/epaa.v22n18.2014

Hwang, S. H., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness, 11*, 559-587. https://doi.org/10.1080/19345747.2018.1496500

Im, M. H., Hughes, J. N., Kwok, O., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology, 51*, 349–365. https://doi.org/10.1016/j.jsp.2013.01.004

Ikeda, M. (2011). PISA in focus: When students repeat grades or are transferred out of School: What does it mean for education systems? http://www.oecd.org/dataoecd/35/58/48363440.pdf

Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies, 2013*, 1–47. https://doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx

Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research, 45*, 613-635. https://doi.org/10.2307/1170067

Jerrim, J., Lopez-Agudo, L. A., & Marcenaro-Gutierrez, O. D. (2022). Grade retention and school entry age in Spain: a structural problem. *Educational Assessment, Evaluation and Accountability, 34*, 331-359. https://doi.org/10.1007/s11092-021-09375-7

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology, 37*, 243-272. https://doi.org/10.1016/S0022-4405(99)00005-9

Jimerson, S. R. (2001). A synthesis of grade retention research: Looking backward and moving forward. *California School Psychologist, 6*, 47–59. https://doi.org/10.1007/BF03340883

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools, 39*, 441-457. https://doi.org/10.1002/pits.10046

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., & Sroufe, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology, 35*, 3-25. https://doi.org/10.1016/S0022-4405(96)00033-7

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22,* 314-339. https://doi.org/10.1037/1045-3830.22.3.314

Justino, D., Pascueiro, L., Franco, L., Santos, R., Almeida, S., & Batista, S. (2014). *Atlas da Educação: Contextos sociais e locais de sucesso e insucesso: Portugal 1991-2012*. CESNOVA. http://www.epis.pt/downloads/mentores/atlas-da-educacao.pdf

Kaplan, A., & Maehr, M. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review, 19*, 141-184. https://doi.org/10.1007/s10648-006-9012-5

Klapproth, F., & Schaltz, P. (2015). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. *European Journal of Psychology of Education, 30*, 119–136. https://doi.org/10.1007/s10212- 014-0232-7

Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences, 50*, 182-194. https://doi.org/10.1016/j.lindif.2016.08.014

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology, 111*, 1432–1446. https://doi.org/10.1037/edu0000353

Lamote, C., Pinxten, M., Van Den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies, 40*, 496–514. http://doi .org/10.1080/03055698.2014.936828

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *Economic Journal, 122,* 208-237. https://doi.org/10.1111/j.1468-0297.2011.02463.x

Le Donné, N. (2014). European variations in socioeconomic inequalities in students' cognitive achievement: The role of educational policies. European Sociological Review, 30(3), 329-343. https://doi.org/10.1093/esr/jcu040

Lubbers, M. J., Van der Werf, M. P. C., Snijders, T. A. B., Creemers, B. P. M., & Kuyper, H. (2006). The impact of peer relations on academic progress in junior high. Journal of School Psychology, 44(6), 491-512. https://doi.org/10.1016/j.jsp.2006.07.005

Ma, X. (2003). Sense of belonging to school: Can schools make a difference? *The Journal of Educational Research, 96*, 340–349. https://doi.org/10.1080/00220670309596617

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology*, 108, 256–273. https://doi.org/10.1037/edu0000059

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S.

(2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109*, 425. doi:10.1037/edu0000144

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*, 319–350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., Trautwein, U., Lüdke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*, 510-524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. Journal of Educational Psychology, 89, 41–54. https://doi.org/10.1037/0022-0663.89.1.41

Martin, A. J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology, 101*, 101-114. https://doi.org/10.1037/a0013100

Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal, 37*, 739–763. https://doi.org/10.1080/01411926.2010.490874

Martin, A. J., Ginns, P., & Papworth, B. A. (2017). Motivation and engagement: Same or different? Does it matter? *Learning and Individual Differences, 55, 150*-162. http://dx.doi.org/10.1016/j.lindif.2017.03.013

Mathys, C., Véronneau, M.-H., & Lecocq, A. (2019). Grade retention at the transition to secondary school: Using propensity score matching to identify consequences on psychosocial adjustment. *Journal of Early Adolescence, 39,* 97–133. https://doi.org/10.1177/0272431617735651

McCoy, A. R., & Reynolds, A. J. (1999). Grade retention and school performance: An extended investigation. *Journal of School Psychology, 37*, 273-298. https://doi.org/10.1016/S0022-4405(99)00012-6

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation and academic achievement. Annal Review of Psychology, 57, 48–503. https://doi.org/10.1146/annurev.psych.56.091103.070258

Mendez, L. M. R., Kim, E. S., Ferron, J., & Woods, B. (2015). Altering school progression through delayed entry or kindergarten retention: Propensity score analysis of long-term outcomes. *Journal of Educational Research, 108*, 186-203. https://doi.org/10.1080/00220671.2013.867474

Ministério da Educação/Direcção Geral de Educação. (n.d.). *Estatísticas do ensino básico e secundário*. https://infoescolas.medu.pt

Mons, N. (2007*). Les nouvelles politiques éducatives* [New educational policies]. Presses Universitaires de France.

Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology,* 104, 603–621. https://doi.org/10.1037/a0027571

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Nascimento, S., & Peixoto, F. (2012). Relações entre o estatuto escolar e o autoconceito, auto-estima e orientações motivacionais em alunos do 9º ano de escolaridade [Relationship between school status, self-esteem and goal orientations among 9th grade students]. *Análise Psicológica, 30*, 421-434. https://doi.org/10.14417/ap.602

National Association of School Psychologists (2021). COVID-19 resources: Guidance on the use of grade retention and special education eligibility to address instructional loss. https://www.nasponline.org/resources-and-publications/resources-and-podcasts/covid-19-resource-center/return-to-school/guidance-on-the-use-of-grade-retention-and-special-education-eligibility-to-address-instructional-loss

Nunes, L. C., Balcão Reis, A., & Seabra, C. (2018). Is retention beneficial to low-achieving students? Evidence from Portugal. *Applied Economics, 50*, 4306–4317. https://doi.org/10.1080/00036846.2018.1444261

Organisation for Economic Co-operation and Development (2003). *Student engagement at school: A sense of belonging and participation: Results from PISA 2000*. OECD Publishing. https://doi.org/10.1787/9789264018938-en

Organisation for Economic Co-operation and Development (2020). *PISA 2018 results: Effective policies, successful schools* (Vol. 5). OECD Publishing. https://doi.org/10.1787/ca768d40-en

Ou, S. R., & Reynolds, A. J. (2010). Grade retention, postsecondary education and public aid receipt. *Educational Evaluation and Policy Analysis, 32*, 118-139. https://doi.org/10.3102/0162373709354334

Özek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. *Education Finance and Policy, 10*(3), 350-377. https://doi.org/10.1162/EDFP_a_00166

Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology, 13*, 297–315. https://doi.org/10.1017/s0954579401002061

Pedraja-Chaparro, F., Santín, D., & Simancas, R. (2015). Determinants of grade retention in

France and Spain: Does birth month matter? *Journal of Policy Modeling, 37*, 820-834. http://dx.doi.org/10.1016/j.jpolmod.2015.04.004

Peixoto, F., & Almeida, L. S. (2010). Self-concept, self-esteem and academic achievement: Strategies for maintaining self-esteem in students experiencing academic failure. *European Journal of Psychology of Education, 25*, 157–175. https://doi.org/10.1007/s10212-010- 0011-z

Peixoto, F., Monteiro, V., Mata, L., Sanches, C., Pipa, J., & Almeida, L. (2016). "To be or not to be retained…That's the question!" Retention, self-esteem, self-concept, achievement goals, and grades. *Frontiers in Psychology, 7*, 1550. https://doi.org/10.3389/fpsyg.2016.01550

Peixoto, F., Pipa, J., Mata, L., Monteiro, V., & Sanches, C. (2017). "To learn, or to be the best?" Achievement goal profiles in pre-adolescents. *Análise Psicológica, 35,* 499-511. https://doi.org/10.14417/ap.1182

Pereira, M. C., & Reis, H. J. (2014). Grade retention during basic education in Portugal: Determinants and impact on student achievement*. Economic Bulletin, 16,* 61–83. https://www.bportugal.pt/sites/default/files/anexos/papers/ab201407_e.pdf

Piaget, J., & Inhelder, B. (1966). *The psychology of the child*. Basic Books.

Pires, L., Santero-Sánchez, R., Macías, C. (2021). School failure in the region of Madrid (Spain): An approximation through diagnostic assessment in 2019. *Sustainability, 13,* 9895. https://doi.org/10.3390/su13179895

Portaria 223-A/2018 (2018). Educação. Diário da República – 1º Suplemento, Série I - No. 149 de 3 de Agosto, pp. 2.23. https://dre.pt/dre/detalhe/portaria/223-a-2018-115886163

Range, B. G., Pijanowski, J., Holt, C. R., & Young, S. (2012). The perceptions of primary grade teachers and elementary principals about the effectiveness of grade-level retention. *Professional Educator, 36*. https://eric.ed.gov/?id=EJ989520

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55. https://doi.org/10.2307/2335942

Rosário, P., Núñez, J. C., Valle, A., González-Pienda, J., & Lourenço, Abílio (2013). Grade level, study time, and grade retention and their effects on motivation, self-regulated learning strategies, and mathematics achievement: a structural equation model. *European Journal of Psychology of Education, 28*, 1311–1331 (2013). https://doi.org/10.1007/s10212-012-0167-9

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*, 169–188. https://doi.org/10.1023/A:1020363010465

Sidali, Z., Pipa, J., & Peixoto, F. (2021, September 1-3). *Determinantes da retenção escolar: Papel das componentes psicossociais* [Poster presentation]. XVI Congresso Internacional Galego-Português de Psicopedagogia, Braga, Portugal.

Santana, M. R. (2019). *Práticas e representações acerca da retenção escolar* [Practices and representations on grade retention] [Doctoral dissertation, Universidade Nova de Lisboa]. Repositório Universidade Nova. https://run.unl.pt/handle/10362/89715

Santos, N. N., Monteiro, V., & Carvalho, C. (2022). Impact of grade retention and school engagement on student intentions to enrol in higher education in Portugal. *European Journal of Education*. https://doi.org/10.1111/ejed.12535

Santos, N. N., Pipa, J., & Monteiro, V. (2023). Analysing grade retention beliefs within teachers' psycho-pedagogic beliefs system. *Teaching and Teacher Education*, *121.* https://doi.org/10.1016/j.tate.2022.103939

Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics, 152*, 154-169. https://doi.org/10.3386/w21509

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407–441. https://doi.org/10.2307/1170010

Shepard, L. A., & Smith, M. L. (1989). *Flunking grades: Research and policies on retention.* Falmer Press. https://doi.org/10.1002/1520-6807(199010)27:4<373::AID-PITS2310270415>3.0.CO;2-A

Skaalvik, E. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology, 89*, 71-81. https://doi.org/10.1037/0022-0663.89.1.71

Slaten, C. D., Ferguson, J. K., Allen, K-A., Brodrick, D-V., & Waters, L. (2016). School belonging: A review of the history, current trends, and future directions. *The Educational and Developmental Psychologist, 33*, 1-15. https://doi.org/10.1017/ edp.2016.6

Sirin, S., & Rogers-Sirin L. (2004). Exploring school engagement of middle-class African American adolescents*. Youth & Society, 35,* 323–340. https://doi.org/10.1177/0044118X03255006

Smith, J., Paquin, S., St-Amand, J., Singh, C., Moreau, D., Bergeron, J., & Leroux, M. (2022). A remediation measure as an alternative to grade retention: A study on achievement motivation. *Psychology in the Schools, 59,* 1209-1221. https://doi.org/10.1002/pits.22677

Smith, M. L., & Shepard, L. A. (1988). Kindergarten readiness and retention: A qualitative

study of teachers' beliefs and practices. American Educational Research Journal, 25, 307–333. https://doi.org/10.2307/1163167

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25,* 1–21 https://doi.org/10.1214/09-STS313

Stuart, E. A., & Green, K. M. (2008). Using full matching to estimate causal effects in nonexperimental studies: Examining the relationship between adolescent marijuana use and adult outcomes. *Developmental Psychology, 44*, 395–406. https://doi.org/10.1037/0012-1649.44.2 .395

Thompson, C. L., & Cunningham, E. K. (2000). *Retention and social promotion: Research and implications for policy*. Eric Digest Number 161. https://files.eric.ed.gov/fulltext/ED449241.pdf

Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal, 29*, 199-223. https://doi.org/10.2307/1162908

Valbuena, J., Mediavilla, M., Choi, A., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys, 35*, 408-451. https://doi.org/10.1111/joes.12406

Van Canegem, T., Van Houtte, M., & Demanet, J. (2021). Grade retention and academic self-concept: A multilevel analysis of the effects of schools' retention composition. *British Educational Research Journal, 47,* 1340-1360. https://doi.org/10.1002/berj.3729

Van Canegem, T., Van Houtte, M., & Demanet, J. (2022a). It's a hard-knock life for us : A multilevel analysis on the association between grade retention and being bullied in 25 countries. *Research Papers in Education*. https://doi.org/10.1080/02671522.2022.2125050

Van Canegem, T., Van Houtte, M., & Demanet, J. (2022b). Grade retention: A pathway to solitude? A cross-national multilevel analysis of the effects of being retained on students' sense of belonging. *Comparative Education Review*. https://doi.org/10.1086/721649

Vandecandelaere, M., Schmitt, E., Vanlaar, G., De Fraine, B., & Van Damme, J. (2016). Effects of kindergarten retention for at-risk children's psychosocial development. *Educational Psychology, 36*, 1354–1389. https://doi.org/10.1080/01443410.2014.950194

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). Time-varying treatments in observational studies: Marginal structural models of the effects of early grade retention on math achievement. *Multivariate Behavioral Research, 51,* 843-864. https://doi.org/10.1080/00273171.2016.1155146

Voelkl, K. E. (1996). Measuring students' identification with school. Educational and *Psychological Measurement, 56,* 760-770. doi:10.1177/0013164496056005003

Vygotsky, L. S. (1986). *Thought and language*. MIT Press.

Wentzel, K. (2002). The contribution of social goal setting to children's school adjustment. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 222-246). Academic Press.

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology, 102*, 135–152. https://doi.org/10.1037/a0016664

Xia, N., & Kirby, S. N. (2009). *Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes*. http://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR678.pdf

Xiang, N., & Chiu, S. (2022). The school matters: Hong Kong secondary schools' grade-retention composition, students' educational performance, and educational inequality. *School Effectiveness and School Improvement.* https://doi.org/10.1080/09243453.2022.2136210

Young, S., Trujillo, N. P., Bruce, M. A., Pollard, T., Jones, J., & Range, B. (2019). Pre- service teachers' views about grade retention as an intervention for struggling students. *Preventing School Failure: Alternative Education for Children and Youth, 63*, 113e120. https://doi.org/10.1080/1045988X.2018.1523124

SECTION II | EMPIRICAL STUDIES

3. STUDY 1: EFFECTIVENESS OF GRADE RETENTION: A SYSTEMATIC REVIEW AND META-ANALYSIS

This chapter is based on Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and a meta-analysis*. Educational Research Review, 34*, 100401. doi:10.1016/j.edurev.2021.100401. Results were also presented at V Encontro Nacional de Jovens Investigadores em Educação (2021), IV Encontro Nacional de Jovens Investigadores em Educação (2020), 18th Biennial EARLI Conference (2019), and XIV Colóquio Internacional de Psicologia e Educação (2018). The first two authors contributed equally for this study and are both considered first author.

Abstract

Research on the effectiveness of grade retention has a long history, yet, has seen an upsurge during the last decade. In this study, we review 84 recent, methodologically sound studies estimating effects of retention in grades K-12 on repeaters' and non-repeaters' development, in a variety of countries across the world, disentangling grade and age comparison results. Based on vote counting analysis and three-level meta-regression analysis we find grade retention to have an average zero effect, indicating that repeaters and non-repeaters seem to show a similar development, on average. At the same time, we find grade retention effects to differ according to some specific effect and study characteristics. More specifically, grade retention seems less effective in countries applying a mixture of grade retention and tracking to tackle student heterogeneity, and when repeaters are compared with non-repeaters of the same age. Conversely, grade retention seems more effective in countries using strategies such as ability grouping, setting, and streaming to deal with student heterogeneity. Positive effects also seem to arise when studying students' psychosocial functioning, when investigating short-run effects, when comparing repeaters with their younger non-retained grade-mates, and when evaluating effects via a regression discontinuity method.

Keywords: grade retention; effectiveness; meta-analysis; K-12 education; learning outcomes

Introduction

Within any given class, there is typically considerable variation in students' academic and psychosocial skills. This heterogeneity challenges educational practitioners and policy makers in their attempts to guarantee high educational standards and raises the question how to deal with it. One can use ability grouping, streaming, tracking, individualized instruction, grade retention, peer tutoring, collaborative learning, or a mixture of practices. The practice we focus on in this study is grade retention. Research on the effectiveness of grade retention has a long history, going back to as early as 1908 (Jackson, 1975), yet, has seen an upsurge since 2000, and especially since 2010. The aim of our study is to review recent, methodologically sound studies estimating effects of retention in grades K-12 on repeaters' and non-repeaters' development (in terms of academic achievement, psychosocial functioning, school career, and job career), in a variety of countries across the world, disentangling grade and age comparison results.

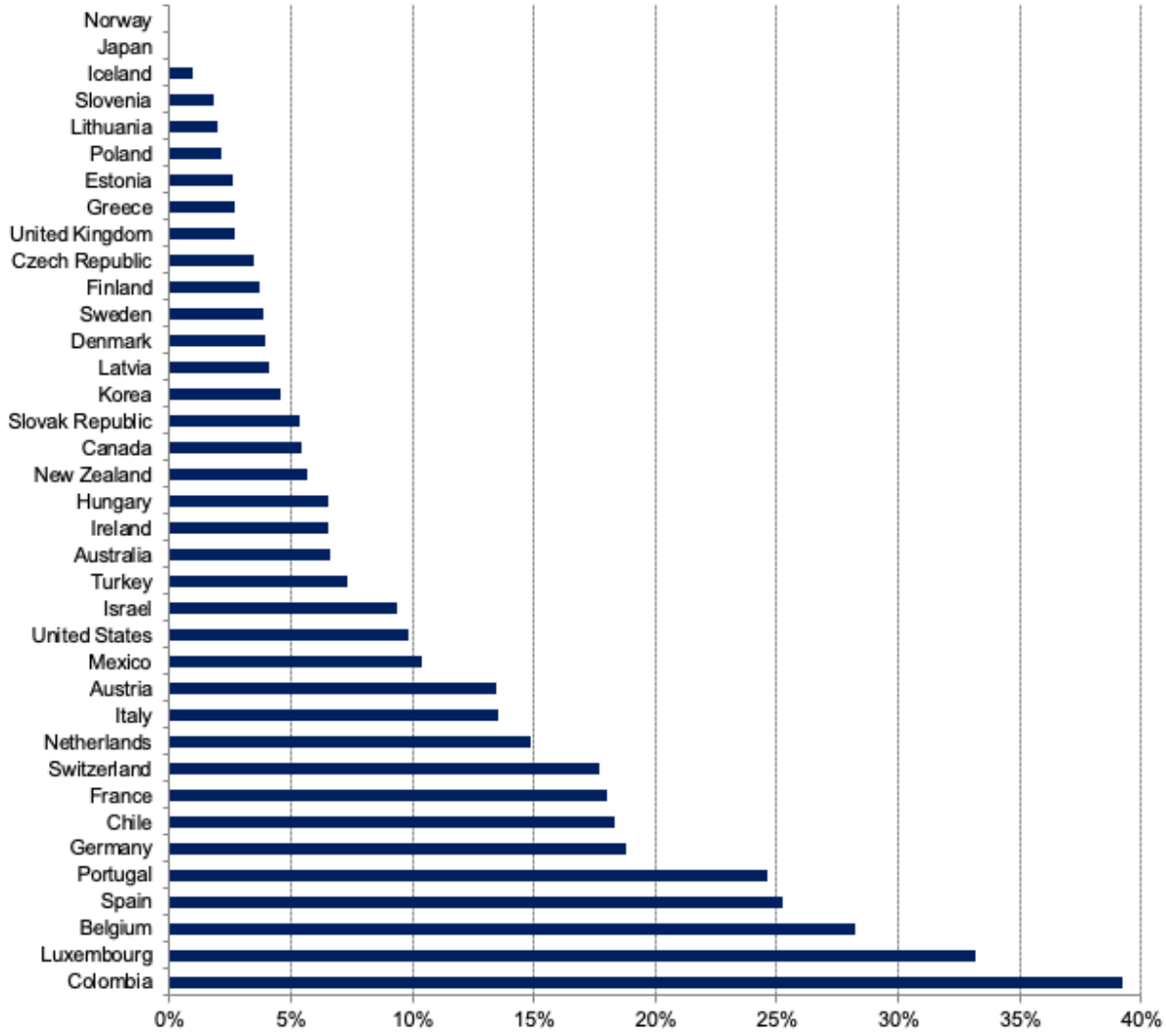
Grade Retention in Practice

Grade retention means that one holds back students who are not keeping up with peers or do not meet a certain predefined level, in the same grade for an extra school year, instead of promoting them to the next grade level (Driessen, 2020). This is one practice that educational practitioners and policy makers can apply to tackle student heterogeneity in grades K-12.

In some countries, grade retention is the most commonly applied practice to tackle student heterogeneity (Dupriez et al., 2008). Mons (2007) refers to this type of countries as 'uniform integration countries'. Examples are France, Spain, and Portugal (see high retention rates in Figure 1). In another set of countries, grade retention is jointly used with ability grouping and tracking to deal with student heterogeneity (Dupriez et al., 2008). Mons (2007) calls this type of countries, among which Germany, Austria, Hungary, Switzerland, Luxembourg, Belgium, and the Netherlands (see high retention rates in Figure 1), 'separation countries'. In some other countries, grade retention is seldom used, and is only used as a last remedy when ability grouping, setting, and streaming appear to insufficiently tackle student heterogeneity (Dupriez et al., 2008). Mons (2007) refers to this type of countries as 'à la carte integration countries'. Examples are the United States, the United Kingdom, Canada, Australia, and New Zealand (see low to moderate retention rates in Figure 1). Finally, in a last set of countries, grade retention is not allowed or rarely applied, because student heterogeneity is mainly dealt with using various forms of differentiation and individualized teaching (Dupriez et al., 2008). Mons (2007) calls this type of countries, among which Denmark, Iceland, Finland, Norway, and Sweden (see low retention rates in Figure 1), 'individualized integration countries'.

Figure 1

Percentage of Students Reporting Having Repeated at Least Once by Age 15, per OECD Member Country, as Derived From the PISA 2018 Student Questionnaire



In addition to these broad application differences across country typologies, the practice of grade retention also differs across countries within a given typology, as well as across states and regions within a given country. Each educational context is unique. Each country, and to some extent also each state and region, has its own educational policy (e.g., regulations related to school entry, presence or absence of high-stakes testing and related regulations concerning grade retention decisions, regulations regarding the application of the retention year as 'mere rehearsal' versus so-called 'retention +' coupling grade retention with many other remedial strategies such as extra tutoring, summer school, etc.), school population, school organization, curriculum, and beliefs about grade retention effectiveness, which altogether influence grade retention application in practice (Agasisti & Cordero, 2017; Demanet & Van Houtte, 2013; Driessen, 2020; Eurydice, 2011; Goos et al., 2013).

Theoretical Debate

As for the beliefs about grade retention effectiveness, educators from 'uniform integration countries', 'separation countries', and 'à la carte integration countries' retain struggling students with the best intentions; intentions that can be linked to diverse theories in education, psychology, and sociology. First of all, many educators are convinced, inspired by maturationist developmental theory (Gesell, 1940; Piaget & Inhelder, 1962), that mastering the learning content related to a specific grade is necessary to succeed in later grades, from a basic 'building blocks' idea, with each grade building upon the previous one. Thus, students who do not yet master specific grade-related knowledge or skills, are in their view best off repeating. The next grade's learning content would be too difficult for them, setting them up for failure and possibly accumulating in a negative school and job career. Second, many educators believe that classes become more homogeneous when the weakest performing students repeat their grade. As a result, they believe that they can instruct their students more efficiently, cover more advanced learning content, and better satisfy students' individual needs, boosting the achievement of all students, both repeaters and non-repeaters (cf. sociocultural developmental theory, Vygotsky, 1978). Third, many educators are convinced that repeaters, during their retention year, start with an advantage in academic knowledge and skills over their new, younger classmates, get the opportunity to refresh and experience new successes, and have accumulated more experience in socializing with peers and coping with conflicts at school in comparison to their new, younger classmates. Therefore, many educators expect grade retention to stimulate repeaters' academic achievement, academic self-concept, self-confidence, pleasure in learning, academic motivation, school well-being, and interpersonal relationships, especially during the retention year (cf. self-efficacy theory, Bandura, 1977; self-determination theory, Deci & Ryan, 2000; social comparison theory, Festinger, 1954). Fourth and finally, many educators believe that the threat of grade retention might act like a 'stick'. By

setting standards, being clear that achievement matters, and enforcing negative consequences if goals are not reached, many educators are convinced that all students (both potential repeaters and non-repeaters) will be motivated to work harder and thus achieve better (cf. self-determination theory, Deci & Ryan).

Holding back struggling students, however, also has potential disadvantages. First, repeating a grade for a big part is simply sheer rehearsal of subjects, while repeaters might already be partly or sufficiently skillful at some subject(s). As a consequence, repeaters – unintendedly – might not receive the continuous intellectual challenges they need, and thus their achievement progress might in fact be impeded instead of stimulated (cf. contextual developmental theory, Bronfenbrenner, 1980; Vygotsky, 1978). Second, repeaters are left behind by their friends, have to get along with new classmates who might bully or reject them, and might feel like 'having failed', seeing all their former classmates advance successfully to the next grade while they have to repeat the grade and its entire subject material. As such, repeating a grade might have a negative impact on repeaters' academic self-concept, self-confidence, pleasure in learning, academic motivation, school well-being, behavior, and interpersonal relationships, especially during the retention year (cf. self-determination theory, Deci & Ryan, 2000; social comparison theory, Festinger, 1954; social control theory, Hirschi, 2002). Third, the threat of grade retention might cause a lot of stress, among both potential repeaters and non-repeaters, possibly impeding students' academic progress (self-determination theory, Deci & Ryan, 2000). Fourth and finally, the practice of grade retention incurs extra costs for the government and the (family of the) repeaters, including the expense of providing all repeaters with an additional year of education and the cost in delaying those students' entry into the labor market by one year. The OECD estimated such costs in Belgium, Spain, the Netherlands, and Portugal to be as high as USD 11,000 or more per repeater (Ikeda & García, 2014).

Previous Research on Grade Retention Effectiveness

Given these disadvantages, do repeaters at least benefit from this practice? This question has been investigated in loads of studies over the last decades, and the results of these studies have in turn been summarized in several meta-analyses (Allen et al., 2009; Bright, 2011; Holmes, 1989; Holmes & Matthews, 1984; Jimerson, 2001) and systematic reviews (Jackson, 1975; Xia & Kirby, 2009). In sum, these meta-analyses and systematic reviews suggest that grade retention is less effective than generally believed by educators, especially in the long run (Driessen, 2020).

These meta-analyses and systematic reviews, however, have five major drawbacks. First, they have almost exclusively focused on grade retention in primary education. This is important, keeping the percentage of primary school repeaters in some countries in mind. Yet,

many students also repeat kindergarten or a grade in secondary education. Effects of grade retention in secondary school might be more negative, for example, because the negative experience of broken peer relationships (tied to grade retention) might be more profound at this age (Ikeda & García, 2014), whereas the opposite may hold for kindergarten retention (Xia & Kirby, 2009, p.30). Somewhat surprisingly, in the meta-analyses and systematic reviews mentioned above (see Allen et al., 2009; Jackson, 1975; Jimerson, 2001; Holmes, 1989; Xia & Kirby, 2009), out of all hundreds of studies summarized, only a small number of studies have assessed effects of repeating kindergarten (28 studies) or a secondary school grade (seven studies).

Second, previous meta-analyses and systematic reviews have predominantly focused on studies regarding grade retention effects in the United States. In the United States, like already pointed out before, student heterogeneity is commonly dealt with using ability grouping, setting, and streaming (Mons, 2007). Having to repeat a grade happens less often (see also Figure 1) and, if applied, is frequently coupled to additional remedial measures during the retention year ('retention +'; Schwerdt et al., 2017). In such case, grade retention might have positive effects on repeaters' academic achievement (because of intensive remediation) and negative effects on repeaters' psychosocial functioning (because of being the sole exception to repeat). As a consequence, results from previous meta-analyses and systematic reviews are not informative for educational practitioners and policy makers from other countries. Out of all hundreds of studies on grade retention effectiveness reviewed so far, only four studies were based on data from students in Canada and Switzerland (see Bright, 2011; Xia & Kirby, 2009), and, surprisingly, none on data from students in any of the countries with the highest number of repeaters.

Third, previous meta-analyses and systematic reviews have mainly summarized findings with regard to repeaters' academic achievement and (to a lesser extent) psychosocial functioning. Insight into these effects is relevant to correctly inform educators and parents who are confronted with a grade retention decision in daily practice. Yet, equally important is knowing how such effects accumulate over time, and influence repeaters' later school and job career (Reschly & Christenson, 2013). Allowing them extra time, do repeaters have a higher chance of obtaining a high school diploma, graduating from tertiary education, having a paid job etc.? These longer-run effects might be completely different from the shorter-run academic and psychosocial effects, yet, have up till now been summarized in one systematic review (Xia & Kirby, 2009) only, based on 18 studies. Similarly, evenly informative is insight into effects of grade retention on repeaters' peers (Reschly & Christenson, 2013). Contrary to repeaters, non-repeaters might experience positive effects of grade retention as classrooms might become more homogeneous and thus easier to instruct, due to retention. Such effects have so far been summarized in no meta-analysis or systematic review at all.

Fourth, previous meta-analyses and systematic reviews have mingled results from grade retention effectiveness studies using a grade approach versus age approach. These two comparison approaches, however, answer very different questions. Grade comparisons compare retainees with their grademates (either their new, younger classmates or their former classmates a school year earlier), and answer the questions how, at the cost of one extra year of education, repeaters develop in comparison to younger students with whom they will eventually graduate and how repeaters affect these younger peers. Age comparisons, by contrast, compare repeaters with their agemates who got promoted to a higher grade, and give an answer to the questions how retainees would have developed and would have affected their peers, had they been promoted to the next grade instead of been held back. Both comparisons and questions are equally relevant for educational practitioners and policy makers, and should never be mingled.

Fifth and finally, previous meta-analyses and systematic reviews have mainly summarized results of methodologically flawed studies. In particular, so far, only three studies reviewed in one meta-analysis (Jackson, 1975) were based on a randomized controlled trial. From an ethical point of view, this is perfectly understandable. It is ethically impossible for a researcher to randomly assign students to an experimental group of grade-retained students versus a control group of grade-promoted students. Therefore, most researchers until now were restricted to following a large cohort of students and registering whom of them were assigned to the experimental group by their teacher(s) and parents, be it based on an a priori defined criterion or not. An important disadvantage of such approach is that under this scenario, retained and promoted students differ on many more characteristics than their retention status alone (e.g., socio-economic status, academic achievement (progress), for an overview, see for instance Xia & Kirby, 2009). Thus, one needs to control for these differences adequately, to not leave open the possibility that these differences rather than grade retention itself cause differences in outcomes to appear. Quasi-experimental methods such as regression discontinuity methods, propensity score methods, instrumental variable methods, difference-in-differences methods, and factor analytic methods are most appropriate for this purpose, yet, have only been used in 19 out of all hundreds of studies reviewed so far (Allen et al., 2009; Bright, 2011; Xia & Kirby, 2009). This has important consequences for the interpretation of findings of previous meta-analyses and systematic reviews, as Allen et al. (2009, p.492) showed that studies using such approaches generally reveal less unfavorable effects on repeaters' academic achievement than do studies tackling selection bias less effectively.

The Present Study

This study aims to summarize the research findings of recent, methodologically sound studies on the effects of retention in grades K-12 on several outcomes, in a variety of countries across the world, disentangling grade and age comparison results. Our research questions are:

RQ1: What is the average effect of grade retention?

RQ2: Does the effect of grade retention differ by: retention timing (i.e., kindergarten, primary education, versus secondary education); retention country heterogeneity management model (i.e., 'separation', 'à la carte integration', versus 'uniform integration'); retention application (i.e., mere rehearsal versus 'retention +'); outcome domain (i.e., academic achievement, psychosocial functioning, school career, versus later job career of repeaters and non-repeaters); outcome timing (i.e., the number of years since the start of the retention year); comparison approach (i.e., grade comparison versus age comparison); and method (i.e., randomized controlled trial, regression discontinuity method, propensity score method, instrumental variable method, difference-in-differences method, versus factor analytic method)?

Method

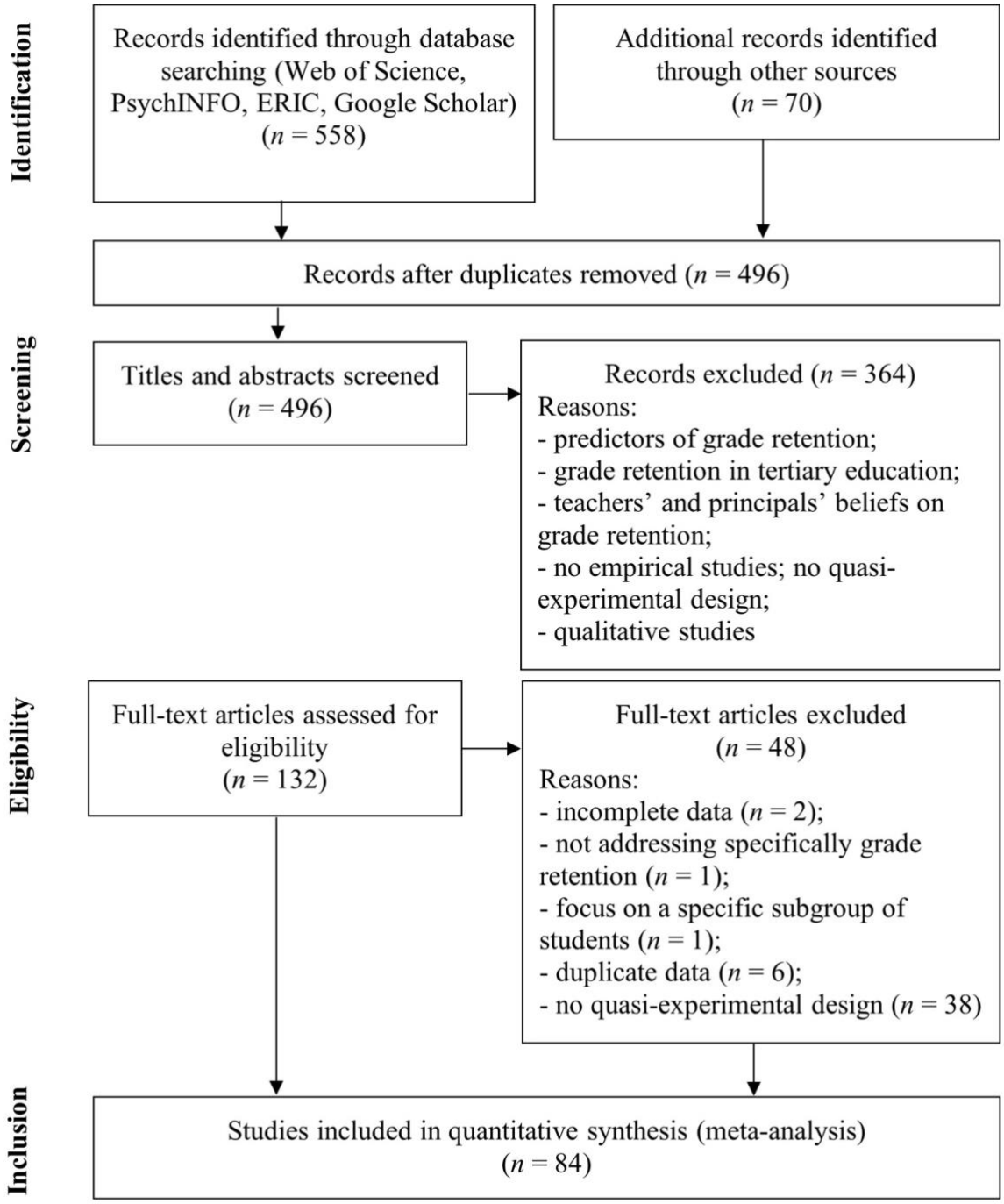
Search of Studies

Studies on grade retention effects were systematically searched through electronic databases (i.e., Web of Science, PsycINFO, and ERIC) and Google Scholar search engine, using the combination of *grade retention, grade repetition, grade failure, repeating a grade, flunking, being held back, repeaters, retainees, or nonpromotion* on the one hand, and *effect** on the other hand (including corresponding translations in French, German, Dutch, Spanish, and Portuguese). In each database, the selected terms were searched either in the title, abstract, or keywords. Additionally, hand searches were conducted based on the reference lists of the retrieved studies and the publication lists of known authors working in the field of grade retention effectiveness. This strategy allowed us to find unpublished studies and grey literature, attenuating publication bias.

In total, 628 studies were identified through electronic databases searches and hand searches (see Figure 2). The studies were exported to EndNote Web software for duplicate removal, and then to Rayann QCRI online tool (Ouzzani et al., 2016) for screening.

Figure 2

Flow Diagram of Study Search and Selection Process Based on the PRISMA Statement (Moher Et Al., 2009)



Selection of Studies

For inclusion in this review, studies had to meet the following five criteria:

- Studies needed to consider grade retention, which was defined as repeating a grade after having spent a full year in that grade. Studies examining retention in grades K-12 were included. Studies on retention in tertiary education were excluded given that repeating a grade in tertiary education has distinctive features and procedures as compared to repeating grades K-12. Studies on school delay (in the broad sense) were also excluded as school delay can have multiple causes in addition to grade retention (e.g., delayed school entrance before kindergarten, placement in developmental kindergarten or a transition room, and enrollment in a grade below age expectations because of foreign newcomer status).
- Studies needed to evaluate all students, not a specific subcategory of students (e.g., second language learners).
- Studies needed to investigate effects of grade retention, not predictors of grade retention. More specifically, studies needed to evaluate effects, among repeaters or non-repeaters, on at least one of the following four outcome domains: (a) academic achievement, (b) psychosocial functioning, (c) school career, and (d) later job career. Academic achievement was operationalized as scores on (standardized) tests in math, languages, or any other domain, grade point averages, and performance as rated by the teacher. Psychosocial functioning was operationalized very broadly, covering academic self-concept, classroom behavior, school well-being, social skills, and relationships with peers, as judged by students themselves, their teachers, their peers, or external observers. School career was operationalized as decisions regarding students' school career such as placement in special education, school mobility, high-school completion, etc., as derived from official student records, class and school lists, or teacher questionnaires. Finally, job career was operationalized as income, unemployment status, social security status, etc., as derived from questionnaires or official governmental records.
- Studies needed to clearly distinguish an experimental group (grade-retained students) and a control group (grade-promoted students), needed to use an experimental design or (by lack of randomization options) any of the following five types of quasi-experimental methods: (a) regression discontinuity methods, (b) propensity score methods, (c) instrumental variable methods, (d) difference-in-differences methods, or (e) factor analytic methods, and needed to provide a reasonable rationale for the assumptions underlying these methods to be credibly met (e.g., robustness checks,

proofs of equivalence of the experimental group and the control group on a considerable amount of observed background characteristics). Studies without a credible control group were excluded. This decision was based on the seminal work of scholars working in the field of causal inference, such as Antonakis, Bendahan, Jacquart, and Lalive (2010), Imbens and Rubin (2015), and Murnane and Willet (2011, p.310), as well as on recommendations by the Cochrane Collaboration (Higgins et al., 2019, pp.321-324), the BEE group (n.d.), and the What Works Clearinghouse (2020).

- Studies needed to be published between January 2000 and October 2019.

The titles and abstracts of all 491 retrieved unique studies were screened based on these five criteria by the first two authors independently, using the blind review feature in Rayann QCRI tool (with 88% agreement). This resulted in a selection of 132 studies for full text reading (see Figure 2), which were again screened by the first two authors independently. Finally, 84 studies covering 1,328 single effects were included in our review (see Figure 2).

Coding of Studies

Following the guidelines from Higgins et al. (2019, p.130), all studies were double-coded. Our coding schema covered the following coding categories: (a) publication information (author, year, and type of publication); (b) general study information (sample size, country, data); (c) grade level in which retention was applied; (d) method; (e) data analysis procedure; (f) outcome examined; (g) grade or age when the outcome was measured; (h) sign of effect; and (i) size of effect.

As for the sign of effect, each single effect was coded taking into consideration its significance and direction (Higgins et al., 2019, p.328), as significantly negative (detrimental), nonsignificant (null), versus significantly positive (beneficial). As for the size of effect, each single effect was coded taking into consideration its magnitude, expressed as Hedges' *g*. First, Cohen's *d* was retrieved or calculated for each single effect, from data directly available in the study (either Cohen's *d*; or *M*, *SD*, and *N* for repeaters and non-repeaters; or *t* and *N* for repeaters and non-repeaters; or *OR*; or η^2) or data available from contacting the study's authors in case data to calculate Cohen's *d* were lacking. Calculations of Cohen's *d* and conversions to Cohens' *d* from other metrics were done based on formulae provided by Borenstein, Hedges, Higgins, and Rothstein (2009, pp.37-49). In case no data in addition to *p*-values above .05 were available from the study or its authors, we assigned each single effect a conservative Cohen's *d* estimate of 0. Second, as recommended by Lakens (2013) and Higgins et al. (2019, p.158), we transformed Cohen's *d* to Hedges' *g*, based on a formula provided by Borenstein et al. (2009, p.39), adjusting our effect size estimates to small sample bias.

As for the outcome examined, it should be noted that, in case of multiple publications based on the same dataset, duplicate outcomes were excluded. For example, if a first publication evaluated students' grade point average (GPA) in grade 3 and a second publication evaluated students' GPA in grades 3-6, grade 3 GPA data were selected from the first publication, and grade 4-6 GPA data from the second publication.

To ensure coding reliability, double-coding was first applied to a subsample of 18 randomly selected studies. Coding results of the two independent coders were confronted to solve discrepancies, until 100% agreement was reached. Second, interrater agreement was calculated for the coded total number of significantly negative, nonsignificant, and significantly positive effects, considering the total number of effects extracted (r = .95) and the coded direction of these effects (Cohen's k = .96). Appendix A1 present the coding results summarized per study.

Synthesis of Studies

In order to estimate the average and differential effects of grade retention based on the 84 studies and 1,328 single effects, we performed a vote counting analysis, a statistical meta-analysis, and a publication bias analysis. All analyses were performed with Excel software and R 4.0.2. software (package metafor, version 2.4-0, Viechtbauer, 2010).

Vote Counting Analysis

We transformed the numbers of significantly negative, nonsignificant, and significantly positive effects per study to proportions per study, and averaged these proportions across studies, to estimate the average effect of grade retention (RQ1). We also performed a sign test on the number of significantly positive effects versus significantly negative effects, to test for significance (Higgins et al., 2019, p.328).

Statistical Meta-analysis

We analyzed all 1,328 single effect sizes by performing a series of multilevel meta-regressions. A multilevel approach was chosen as most studies (94%) yielded multiple effect sizes, violating the assumption of independence of effect sizes underlying a traditional fixed-effects approach. More specifically, we fitted a series of three-level linear regression models. A three-step sequential modeling approach was used, in which each successive model reflected an increased complexity. In a first step (Model 1), a regression model without any predictors was fitted in order to estimate the average effect of grade retention (RQ1):

Level 1: $ES_{ijk} = \beta_{0jk} + e_{ijk}$ with $e_{ijk} \sim N(0, \quad \sigma_e^2)$

Level 2: $\beta_{0jk} = \pi_{00k} + v_{0jk}$ with $v_{0jk} \sim N(0, \quad \sigma_v^2)$

Level 3: $\pi_{00k} = \gamma_{000} + u_{00k}$ with $u_{00k} \sim N(0, \quad \sigma_u^2)$

where e_{ijk} equals the sampling variation, v_{0jk} the within-study variation, u_{00k} the between-study variation (estimated to be 0.06%, 48.47%, and 51.47%, respectively) and γ_{000} the average size of effect of grade retention. In a second step (Models 2a to 2g), all effect-level predictors (i.e., outcome domain, outcome timing, comparison approach, method) and study-level predictors (i.e., retention timing, retention country heterogeneity management model, and retention application) were added to the fixed part of the baseline model, one by one, in order to estimate potential differential effects of grade retention (RQ2). In a third and final step (Model 3), all significant predictors from the second step were added together to the fixed part of the model, in order to estimate joint moderating effects (RQ2). Importantly, we estimated all parameters using the restricted maximum likelihood (REML) procedure, and we adjusted the confidence intervals for ES_{ijk} using the Knapp-Hartung (KH) method. We carried out *F*- and *t*-tests, to test for significance.

Publication Bias Analysis

We checked for publication bias in our sample via two different approaches. First, we used the small sample bias method (Borenstein et al., 2011, pp.277-292; Fernandez-Castilla, et al., 2018; Higgins et al., 2019, pp.362-369). We plotted a funnel plot, with all 1,328 effect sizes plotted on the horizontal axis and their standard errors on the vertical axis (see Appendix A3), and we applied Egger's weighted regression test to evaluate funnel plot asymmetry. The results indicated an absence of publication bias (z = 0.18, p = .680). Second, we applied an additional moderator regression analysis (Model 2h), similar in approach as those discussed in the previous paragraph (Models 2a to 2g), but now with publication status as study-level dummy predictor added to the fixed part of the baseline model. Again, the results suggested an absence of publication bias, by revealing no differences in grade retention effect sizes between published and unpublished studies (F(1, 1326) = 0.10, p = .680).

Results

As can be seen in Appendix A1, most studies focused on grade retention in primary education ($k_{studies}$ = 40), were conducted in 'à la carte integration countries' ($k_{studies}$ = 42, which in our case are studies from the United States only), evaluated effects on repeaters' academic achievement ($k_{studies}$ = 34) or a combination of outcomes ($k_{studies}$ = 29), used either a grade comparison approach ($k_{studies}$ = 32) or an age comparison approach ($k_{studies}$ = 43), not both ($k_{studies}$ = 9), used a propensity score method ($k_{studies}$ = 43), followed students for at least three school years ($k_{studies}$ = 67; M_{duration} = 6.93 years), examined at least 1,000 students ($k_{studies}$ = 61; range from 52 students to more than 1,000,000 students), were published in a journal ($k_{studies}$

= 64), and were so far not examined in the abovementioned, previously published meta-analyses and systematic reviews ($k_{studies}$ = 72).

Average Effect of Grade Retention

On average, 35% of effects were significantly negative, 41% non-significant, and 24% significantly positive. Significantly negative effects (favoring non-repeaters) were as present as significantly positive effects (favoring repeaters) according to the vote count sign test, z = -0.60, p = .550. The mean effect size estimated in the three-level meta-regression baseline model (Model 1) was -0.04 (SE = 0.04, 95% CI = [-0.11**;** 0.03], t(1327) = -1.11, p = 0.266), indicating that repeaters and non-repeaters, on average, show a similar development, pooling all outcomes we investigated. The forest plot, showing the effect sizes per study, is presented in Appendix A2.

Differential Effects of Grade Retention

The differential effect sizes estimated in the three-level meta-regression models with predictors (Models 2a to 2g) are summarized in Table 1. Effect sizes were estimated to differ, on average, by retention country heterogeneity management model (F(2, 1294) = 3.37, p = .035), retention application (F(1, 1326) = 11.57, p < .001), outcome domain (F(4, 1323) = 5.99, p < .001), outcome timing (F(1, 1326) = 9.43, p = .002), comparison approach (F(1, 1326) = 23.60, p < .001), and method (F(5, 1322) = 3.06, p = .009), but not retention timing (F(2, 1325) = 1.46, p = .232). Pairwise t-tests (with dummy variables instead of categorical variables) showed that effect sizes and thus effects of grade retention were significantly more negative (a) in ‘separation countries’ (Hedges’ g = -0.18), (b) when retention was applied as ‘mere rehearsal’ (Hedges’ g = -0.10), and (c) when applying an age comparison approach (Hedges’ g = -0.11). By contrast, pairwise t-tests also showed that effect sizes and thus effects of grade retention were significantly more positive (a) in ‘à la carte integration countries’ (Hedges’ g = 0.05), (b) when retention was applied as ‘retention +’ (Hedges’ g = 0.20), (c) when looking at effects on repeaters’ psychosocial functioning (Hedges’ g = 0.08), (d) when applying a grade comparison approach (Hedges’ g = 0.05), and (e) when testing effects with a regression discontinuity method (Hedges’ g = 0.17), and an omnibus F-test showed that effect sizes and thus effects of grade retention were significantly more positive when looking at effects in the shorter run (with Hedges’ g = 0.01 in the retention year, diminishing with 0.01 each year afterwards). Differential effect sizes by retention country heterogeneity management model (t(1290) = 2.14, p = .033), outcome timing (t(1290) = -3.30, p = .001), and comparison approach (t(1290) = -4.99, p < .001) were estimated to keep their significance in a three-level meta-regression model with joint predictors (Model 3, see Table 1), proving their robustness.

Table 1

Results of Three-Level Moderator Analyses of Grade Retention Effects

						Models 2a to 2h		Model 3	
Moderator	$k_{studies}$[a]	k_{ES}	ES (*g*)	95% CI		Test statistic	*p*	Test statistic	*p*
Retention timing						*F*(2,1325) = 1.46	.232	-	-
- Kindergarten	14	177	-0.13	-0.27	0.01				
- Primary education	51	794	-0.02	-0.09	0.06				
- Secondary education	28	357	-0.04	-0.13	0.05				
Retention country heterogeneity management model						*F*(2,1294) = 3.37	.035	*t*(1290) = 2.14	.033
- Separation	20	322	-0.18	-0.32	-0.03				
- Uniform integration	20	163	-0.08	-0.23	0.07				
- À la carte integration	42	809	0.05	-0.05	0.14				
Retention application						*F*(1,1326)= 11.57	< .001	*t*(1290) = 1.68	.094
- Mere rehearsal	69	789	-0.10	-0.17	-0.02				
- Retention +	15	539	0.20	0.04	0.35				
Outcome domain						*F*(4,1323) = 5.99	< .001	*t*(1290) = 1.26	.208
- Repeaters' academic achievement	58	524	-0.07	-0.14	0.01				
- Repeaters' psychosocial functioning	22	326	0.08	-0.01	0.17				
- Repeaters' school career	29	446	-0.10	-0.19	-0.01				
- Repeaters' job career	2	10	-0.07	-0.34	0.19				
- Non-repeaters' development[b]	4	22	0.02	-0.26	0.30				
Outcome timing	84	1328				*F*(1, 1326) = 9.43	.002	*t*(1290) = -3.30	.001
- Retention year			0.01	-0.07	0.08				
Comparison approach						*F*(1,1326)= 23.60	< .001	*t*(1290) = -4.99	< .001
- Grade	41	559	0.05	-0.03	0.12				
- Age	52	769	-0.11	-0.18	-0.04				
Method[c]						*F*(5, 1322) = 3.06	.009	*t*(1290) = 1.17	.242
- DID	9	76	-0.05	-0.20	0.10				
- FAD	3	29	-0.06	-0.43	0.32				
- IV	15	74	-0.15	-0.31	0.00				
- PS	46	591	-0.09	-0.18	0.00				
- RCT	1	17	0.06	-0.56	0.69				
- RD	17	541	0.17	0.03	0.31				
Publication status						F(1, 1326) = 0.10	.753	-	-
- Not peer reviewed	20	184	-0.06	-0.21	0.09				
- Peer reviewed	64	1144	0.03	-0.12	0.05				

Note. [a]A sensitivity analysis was carried out to check the effect of leaving out subgroups with less than 10 studies. Results (available upon request from the authors) equal those presented here. [b]Effects regarding non-repeaters' academic achievement, psychosocial functioning, school career, and job career (retrieved from 4 studies) were collapsed into one category due to small sample size.[c] DID = differences-in-differences method; FAD = factor analytic method; IV = instrumental variable method; PSM = propensity score method; RCT = random control trial; RD = regression discontinuity method.

Discussion

Summary of the Findings

A large body of research has examined the effectiveness of holding students back in grade. We reviewed 84 recent studies on this topic. Three important findings emerge from our study. First, our study seems to indicate that the field of grade retention effectiveness research has considerably improved during the last two decades. More attention has been paid to examining effects of grade retention in a variety of countries across the world, and especially in countries where grade retention is most frequently applied (39 of the 84 reviewed studies were conducted in 'separation countries' and 'uniform integration countries', whereas almost none in the past). In addition, more attention has been paid to evaluating effects of grade retention on a broad spectrum of outcomes, instead of academic achievement alone (29 of the 84 reviewed studies looked at a combination of outcomes). And finally, more attention has been paid to evaluating grade retention effects using a range of sound methods, among which a propensity score method currently seems to be used most often, especially among scholars in education and psychology.

Second, our study seemingly shows that, on average, grade retention has no (negative nor positive) effect on repeaters' and non-repeaters' development. We find as many significantly negative effects as significantly positive effects, and an average effect size of -0.04. This seems contrary to all previously published meta-analyses and systematic reviews that suggested grade retention to be less effective than generally believed by educators, especially in the long run (Allen et al., 2009; Bright, 2011; Holmes, 1989; Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Xia & Kirby, 2009). It should be kept in mind, however, that we selected studies using more rigorous methods. Allen et al. (2009) obtained a similar zero average effect size when restricting the study sample to high-quality studies.

Third, regardless of an average zero effect, our study at the same time seems to indicate differential effects to be present, under specific circumstances. More specifically, we find our effect sizes to range from -0.18 to 0.20. Grade retention seems least effective in Belgium, Germany, Luxembourg, and Uganda ('separation countries') and most effective in the United States, especially under the 'retention +' regime, nowadays in effect in many US regions. This finding is in line with contextual developmental theory (Bronfenbrenner, 1980; Vygotsky, 1978) stating that students need continuous intellectual challenges, and not sheer rehearsal of subjects, while already being partly or sufficiently skillful at some subject(s).

Furthermore, of all outcome domains investigated, repeaters' psychosocial functioning seems most positively affected by grade retention. This finding is congruent with self-efficacy theory (Bandura, 1977), self-determination theory (Deci & Ryan, 2000), and social comparison theory (Festinger, 1954), altogether predicting that repeaters (as compared to younger non-retained grademates) will show a more positive academic self-concept, more self-confidence,

more pleasure in learning, a better academic motivation, a higher school well-being, and more positive interpersonal relationships, especially during the retention year, due to having advantages in academic knowledge and skills at the start of the retention year, having more experience in socializing with peers and coping with conflicts, and getting chances to 'refresh' and experience new successes throughout the retention year.

Repeaters' school career and later job career (both long-run effects) seem most negatively affected by grade retention. Compared with similar non-repeaters, repeaters have a higher chance of placement in special education, track downgrading, school absences, school suspensions, school drop-out, non-entry into tertiary education, low income, public aid receipt, etc. Apparently, very small negative effects on repeaters' academic achievement seem to accumulate with time, creating a negative 'snowball effect' on repeaters' school and job career, a finding in line with that found by Xia and Kirby (2009).

Next, grade retention seems most effective under a grade comparison approach and least effective under an age comparison approach. This finding is congruent with findings of Allen et al. (2009) and Xia and Kirby (2009), and emphasizes, once more, the importance of disentangling grade and age comparison results.

Finally, grade retention effects seem most positive when evaluated via a regression discontinuity method. This finding is not surprising as this method selects 'marginal' students (i.e., students scoring just below or above a certain predefined achievement level), and thus disregards the lowest achievers who need a wider range of remediation measures. This finding stresses the need for educational practitioners and policy makers to interpret results of grade retention effectiveness studies with care, always keeping in mind the method used.

Study Limitations and Suggestions for Future Research

Although the results of this study provide insights into the effects of grade retention, several limitations should be acknowledged. First, looking at our selection procedure, we excluded a large share of studies because they did not clearly distinguish an experimental versus control group, nor used an experimental design or (by lack of randomization options) quasi-experimental method, nor provided a reasonable rationale for the assumptions underlying these methods to be credibly met (i.e., inclusion criterion 4; cf. recommendations by Antonakis et al., 2010; Higgins et al., 2019; Imbens & Rubin, 2015; Murnane & Willet, 2011, the BEE group, and the What Works Clearinghouse). Including such studies (e.g., observational studies with regression-based analyses) would have allowed us to test the differential effect of grade retention in yet another method subcategory (in addition to a randomized controlled trial, regression discontinuity method, etc.). Based on findings of Allen et al. (2009, p.492) focusing on grade retention effects on repeaters' academic achievement, we could hypothesize that the grade retention effect size among such studies would be 0.22

lower (and thus negative), and that the average grade retention effect size would drop, as a result of including such studies in our meta-analysis sample. Future research could confirm this hypothesis. We decided not to include such studies, for reasons of methodological rigor.

In addition, after applying our five inclusion criteria altogether, 84 studies remained. It appeared that these 84 studies investigated grade retention effects in a still rather fragmentary way: (like older studies reviewed in previous meta-analyses and systematic reviews) they still less frequently focus on grade retention in kindergarten and secondary education, still seldom evaluate effects on repeaters' job career and non-repeaters' development, and still are not or rarely conducted in countries such as Colombia, Luxembourg, Spain, or Portugal, four countries in the OECD top 5 regarding grade retention rates, nor countries such as Finland or Sweden, two countries in the OECD bottom 10 regarding grade retention rates. To better generalize and contextualize our findings, we believe more research is needed considering a variety of retention timings, outcomes, and contexts.

Second, looking at our coding procedure, we coded our retention application moderator as 'mere rehearsal' versus 'retention +', which is a very raw categorization, based on general educational policy regulations. In practice, repeaters might get additional support, by different people, and probably for different reasons. Research on this matter is almost absent (Reschly & Christenson, 2013), but highly needed to better understand what happens during the retention year and how this shapes retention effects.

Third, looking at our meta-analysis procedure, we were able to explore seven potentially important moderators of grade retention effects: (a) retention timing, (b) retention country heterogeneity management model, (c) retention application, (d) outcome domain, (e) outcome timing, (f) comparison approach, and (g) method. Unfortunately, we could not add student characteristics such as ability or second language learning status as an extra set of moderators, because fewer studies so far examined such differential effects and our meta-analysis sample as a result would be too small to test for significance. With more research forthcoming in the next decade, however, our exploration could be extended in such way, resulting in a better understanding of who does (or does not) benefit from grade retention, highly relevant questions raised by educational policy makers and practitioners.

Fourth, looking at our publication bias procedure, we were constrained in detecting bias via different methods, simply because such methods so far have been hardly developed nor tested in the context of multilevel meta-analysis (Fernandez-Castilla et al., 2018). This seems an important venue for future research.

Implications for Educational Policy and Practice

Our study revealed that repeaters seem worst off in 'separation countries', applying grade retention and tracking to tackle student heterogeneity, and best off in the United States,

applying ability grouping, setting, and streaming (and to a much lesser extent grade retention) to deal with student heterogeneity. Keeping in mind the costs grade retention brings about to the economy and the (family of the) individual student, we advocate countries, and particularly 'separation countries', efforts to make available more resources for schools (a) to identify at an earlier stage those students at risk for developing school difficulties, by means of a more regular, careful monitoring of their students' progress (especially those countries not yet doing so), (b) to provide *all* at-risk students with targeted remedial interventions that have been proven to be cost-effective (e.g., self-regulation interventions, reading interventions, 1-to-1 or small-group tuition, summer schools; Baye et al., 2019; Higgins et al., 2016; Slavin et al., 2011), be it inside or outside the classroom, from the exact first moment difficulties arise, and (c) to pay considerable attention to repeaters and to give them all services needed, if it is decided to hold particular students back in grade, in their best interest, whatever the reason may be (e.g., outlining a specific plan highlighting how the retention year will be different from the previous year, for example, with regard to level of intensity and duration of support).

References

References marked with an asterisk indicate studies included in the systematic review.

* Afsa, C. (2011). *Faut-il supprimer le redoublement? Document de travail, série «Études», N°2011-E05*. http://www.education.gouv.fr/archives/2012/refondonslecole/wp-content/uploads/2012/07/fiche_thematique_depp_n_7_la_question_du_redoublement_2012.pdf

Agasisti, T., & Cordero, J. M. (2017). The determinants of repetition rates in Europe: Early skills or subsequent parents' help? *Journal of Policy Modeling, 39*(1), 129-146. https://doi.org/10.1016/j.jpolmod.2016.07.002

* Alet, É., Bonnal, L., & Favard, P. (2013). Repetition: Medicine for a short-run remission. *Annals of Economics and Statistics, 111*, 227-250. https://doi.org/10.2307/23646332

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis, 31*(4), 480-499. https://doi.org/10.3102/0162373709352239

* Allensworth, E. M. (2005). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Educational Evaluation and Policy Analysis, 27*(4), 341-364. https://doi.org/10.3102/01623737027004341

* André, P. (2009). *Is grade repetition one of the causes of early school dropout? Evidence from Senegalese primary schools*. Retrieved from https://mpra.ub.uni-muenchen.de/25665/3/MPRA_paper_25665.pdf

* Andrew, M. (2014). The scarring effects of primary-grade retention? A study of cumulative advantage in the educational career. *Social Forces, 93*(2), 653-685. https://doi.org/10.1093/sf/sou074

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*(6), 1086-1120. https://doi.org/10.1016/j.leaqua.2010.10.010

* Appelmans, S. (2016). *Secundaire schoolloopbanen van zittenblijvers in het eerste leerjaar van het lager onderwijs [Secondary school careers of first-grade repeaters]* [Unpublished Master´s thesis]. KULeuven

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

* Battistin, E., & Schizzerotto, A. (2012). *Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in Italy*. https://www.econstor.eu/bitstream/10419/69377/1/73381073X.pdf

* Battistin, E., & Schizzerotto, A. (2019). Threat of grade retention, remedial education and student achievement: Evidence from upper secondary schools in Italy. *Empirical Economics, 56*(2), 651-678. https://doi.org/10.1007/s00181-018-1443-6

Baye, A., Inns, A., Lake, C., & Slavin, R. E. (2019). A synthesis of quantitative research on reading programs for secondary students. *Reading Research Quarterly, 54*(2), 133-166. https://doi.org/10.1002/rrq.229

* Belot, M., & Vandenberghe, V. (2014). Evaluating the "threat" effects of grade repetition: Exploiting the 2001 reform by the French-speaking community of Belgium. *Education Economics, 22*(1), 73-89. https://doi.org/10.1080/09645292.2011.607266

The Best Evidence Encyclopedia group (n.d.). *Criteria for inclusion in the Best Evidence Encyclopedia*. http://www.bestevidence.org/methods/criteria.htm

* Bhattacharya, S. (2008). *The effect of grade retention on child test scores [*Paper presentation]. 58th annual meeting of the American Economic Association, New Orleans, LA. http://www.aeaweb.org/annual_mtg_papers/2008/2008_445.pdf

* Bietenbeck, J. (2014). *Learning from adversity? Short- and long-term spillover effects from grade retention in kindergarten*. https://www.researchgate.net/publication/268447733_Learning_from_Adversity_Short-and_Long-Term_Spillover_Effects_from_Grade_Retention_in_Kindergarten

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley.

Bright, A. D. (2011). *A meta-analysis of the effects of grade retention of K-6 students on student achievement: 1990-2010.* [Doctoral dissertation, University of Alabama]. http://acumen.lib.ua.edu/content/u0015/0000001/0000747/u0015_0000001_0000747

Bronfenbrenner, U. (1980). *The ecology of human development: Experiments by nature and design*. Harvard University Press.

* Cham, H. N., Hughes, J. N., West, S. G., & Im, M. H. (2015). Effect of retention in elementary grades on grade 9 motivation for educational attainment. *Journal of School Psychology, 53*(1), 7-24. https://doi.org/10.1016/j.jsp.2014.10.001

* Chen, X. X., Liu, C. F., Zhang, L. X., Shi, Y. J., & Rozelle, S. (2010). Does taking one step back get you two steps forward? Grade retention and school performance in poor areas in rural China. *International Journal of Educational Development, 30*(6), 544-559. https://doi.org/10.1016/j.ijedudev

* Cockx, B., Picchio, M., & Baert, S. (2019). Modeling the effects of grade retention in high school. *Journal of Applied Econometrics, 34*(3), 403-424. https://doi.org/10.1002/jae.2670

* Cooley-Fruehwirth, J. C., Navarro, S., & Takahashi, Y. (2016). How the timing of grade retention affects outcomes: Identification and estimation of time-varying treatment effects. *Journal of Labor Economics, 34*(4), 979-1021. https://doi.org/10.1086/686262

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement*, *24*(4), 417-434. https://doi.org/10.1080/09243453.2012.727834

* Detienne, L. (2017). *Zittenblijven in het eerste jaar van het secundair onderwijs en het effect op het schoolwelbevinden [Effects of retention in grade 7 on school wellbeing].* [Unpublished master's thesis]. KULeuven.

* d'Haultfoeuille, X. (2010). A new instrumental method for dealing with endogenous selection. *Journal of Econometrics, 154*(1), 1-15. https://doi.org/10.1016/j.jeconom.2009.06.005

* Diris, R. (2012). *Chapter 7: The impact of grade retention: An IV approach to estimating heterogeneous treatment effects.* [Unpublished doctoral dissertation]. Maastricht University.

* Diris, R. (2017). Don't hold back? The effect of grade retention on student achievement. *Education Finance and Policy, 12*(3), 312-341. https://doi.org/10.1162/EDFP_a_00203

* Dong, Y. Y. (2010). Kept back to get ahead? Kindergarten retention and academic performance. *European Economic Review, 54*(2), 219-236. https://doi.org/10.1016/j.euroecorev

Driessen, G. (2020). *Grade retention, grade repetition, holding back a grade*. https://encyclopedia.pub/item/revision/bcdab43e3cb9673d0d197cef8a9b4c18

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity?. *Comparative Education Review*, *52*(2), 245-273. https://doi.org/10.1086/528764

* Ehmke, T., Drechsel, B., & Carstensen, C. H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. *Studies in Educational Evaluation, 36*(1), 27-35. https://doi.org/10.1016/j.stueduc.2010.10.003

* Ehmke, T., Drechsel, B., & Carstensen, C. H. (2013, August). *Explaining the learning progress of retained students and low-achieving students in mathematics* [Paper presentation]. Biennial meeting of the European Association for Research on Learning and Instruction (EARLI), Munich, Germany.

* Ehmke, T., Salzer, C., Pietsch, M., Drechsel, B., & Muller, K. (2017). Competence development in the school year after PISA 2012: Effects of grade retention. *Zeitschrift Fur Erziehungswissenschaft, 20*, 99-124. https://doi.org/10.1007/s11618-017-0752-4

* Eide, E. R., & Showalter, M. H. (2001). The effect of grade retention on educational and labor market outcomes. *Economics of Education Review, 20*(6), 563-576. https://doi.org/10.1016/S0272-7757(00)00041-8

* Eren, O., Depew, B., & Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics, 153*, 9-31. https://doi.org/10.1016/j.jpubeco.2017.07.002

Eurydice. (2011). *Grade retention during compulsory education in Europe: Regulations and statistics* (9789292011406). https://op.europa.eu/en/publication-detail/-/publication/10f14860-12a3-4f9e-b10a-32bee875420d

Fernandez-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2018). Detecting selection bias in meta-analyses with multiple outcomes: A simulation study. *Journal of Experimental Education*. https://doi.org/10.1080/00220973.2019.1582470

* Ferreira Sequeda, M., Golsteyn, B. H., & Parra-Cely, S. (2018). *The effect of grade retention on secondary school performance: Evidence from a natural experiment* [IZA discussion paper No. 1160]. https://www.econstor.eu/bitstream/10419/185064/1/dp11604.pdf

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*(2), 117-140. https://doi.org/10.1177/001872675400700202

* Garcia-Perez, J. I., Hidalgo-Hidalgo, M., & Robles-Zurita, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied Economics, 46*(12), 1373-1392. https://doi.org/10.1080/00036846.2013.872761

* Gary-Bobo, R. J., Gousse, M., & Robin, J. M. (2016). Grade retention and unobserved heterogeneity. *Quantitative Economics, 7*(3), 781-820. https://doi.org/10.3982/QE524

* Geng, T., & Rockoff, J. (2017). *Does repeating a grade make students (and parents) happier? Regression discontinuity evidence from New York City*. https://www.udesa.edu.ar/sites/default/files/rockoff.pdf

Gesell, A. (1940). *The first five years of life*. Harper.

* Glick, P., & Sahn, D. E. (2010). Early academic performance, grade repetition, and school attainment in Senegal: A panel data analysis. *World Bank Economic Review, 24*(1), 93-120. https://doi.org/10.1093/wber/lhp023

Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review, 57*(1), 54. https://doi.org/10.1086/667655

* Goos, M., Van Damme, J., Onghena, P., & Petry, K. (2011, March). *First-grade retention: Effects on children's actual and perceived performance throughout elementary education*. Paper presented at the 1st spring conference of the Society for Research on

Educational Effectiveness, Washington, DC. http://www.sree.org/conferences/2011/program/downloads/abstracts/107.pdf

* Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*(3), 323-347. https://doi.org/10.1016/j.jsp.2013.03.002

* Greene, J. P., & Winters, M. A. (2007). Revisiting grade retention: An evaluation of Florida's test-based promotion policy. *Education Finance and Policy, 2*(4), 319-340. https://doi.org/10.1162/edfp.2007.2.4.319

* Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy: Who is retained? Who benefits academically? *Economics of Education Review, 28*(1), 135-142. https://doi.org/10.1016/j.econedurev.2008.02.002

* Hamlin, C. (2013). *The practice and effects of a school district's retention policies.* [Doctoral dissertation, University of Southern California]. https://search.proquest.com/openview/20b2b8723ca2f8e53af5d88fabc17251/1?pq-origsite=gscholar&cbl=18750&diss=y

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analysis relating to achievement*. Routledge.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Higgins, S., Katsipataki, M., Villanueva-Aguilera, A. B., Coleman, R., Henderson, P., Major, L. E., ... & Mason, D. (2016). *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit*. https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/

* Hill, J., Weiss, C., & Zhai, F. H. (2011). Challenges with propensity score strategies in a high-dimensional setting and a potential alternative. *Multivariate Behavioral Research, 46*(3), 477-513. https://doi.org/10.1080/00273171.2011.570161

Hirschi, T. (2002). *Causes of delinquency*. New Brunswick, NJ: Transaction.

* Hofer, K. G., Farran, D., Lipsey, M., Aydogan, C., & Bilbrey, C. (2010). *Using propensity scores to estimate the effect of early grade retention* [Paper presentation]. Seventh biennial Conference on Research Innovations in Early Intervention (CRIEI), San Diego, CA.

Holmes, C. T. (1989). Grade-level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp.16-33). The Falmer Press.

Holmes, C. T., & Matthews, K. M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*(2), 225-236. https://doi.org/10.3102/00346543054002225

* Hong, G. L., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*(3), 205-224. https://doi.org/10.3102/01623737027003205

* Hong, G. L., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*(475), 901-910. https://doi.org/10.1198/016214506000000447

* Hong, G. L., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis, 29*(4), 239-261. https://doi.org/10.3102/0162373707309073

* Hong, G. L., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44*(2), 407-421. https://doi.org/10.1037/0012-1649.44.2.407

* Hughes, J. N., Cao, Q., West, S. G., Smith, P. A., & Cerda, C. (2017). Effect of retention in elementary grades on dropping out of school early. *Journal of School Psychology, 65*, 11-27. https://doi.org/10.1016/j.jsp.2017.06.003

* Hughes, J. N., Chen, Q., Thoemmes, F., & Kwok, O. M. (2010). An investigation of the relationship between retention in first grade and performance on high stakes tests in third grade. *Educational Evaluation and Policy Analysis, 32*(2), 166-182. https://doi.org/10.3102/0162373710367682

* Hughes, J. N., West, S. G., Kim, H., & Bauer, S. S. (2017). Effect of early grade retention on school completion: A prospective study. *Journal of Educational Psychology, 110*(7), 974–991. https://doi.org/10.1037/edu0000243

* Hwang, S. H., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness, 11*(4), 559-587. https://doi.org/10.1080/19345747.2018.1496500

Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies, 2013*(1), 1-47. https://doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx

* Im, M. H., Hughes, J. N., Kwok, O., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology, 51*(3), 349-365. https://doi.org/10.1016/j.jsp.2013.01.004

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences: An introduction*. Cambridge University Press.

Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research, 45*(4), 613-635. https://doi.org/10.2307/1170067

* Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics, 86*(1), 226-244. https://doi.org/10.1162/003465304323023778

* Jacob, B. A., & Lefgren, L. (2009). The effect of grade retention on high school completion. *American Economic Journal: Applied Economics, 1*(3), 33-58. https://doi.org/10.1257/app.1.3.33

Jimerson, S. R. (2001). A synthesis of grade retention research: Looking backward and moving forward. *The California School Psychologist, 6*, 47-59. https://doi.org/10.1007/BF03340883

* Keslair, F. (2007). *Le redoublement permet-il d'améliorer les résultats scolaires? Le cas d'un redoublement au début de l'école primaire [Unpublised* Master's thesis]. Paris School of Economics.

* Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences, 50*, 182-194. https://doi.org/10.1016/j.lindif.2016.08.014

* Koppensteiner, M. F. (2014). Automatic grade promotion and student performance: Evidence from Brazil. *Journal of Development Economics, 107*, 277-290. https://doi.org/10.1016/j.jdeveco.2013.12.007

* Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology, 111*(8), 1432-1446. https://doi.org/10.1037/edu0000353

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863

* Lamote, C., Pinxten, M., Van den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies, 40*(5), 496-514. https://doi.org/10.1080/03055698.2014.936828

* Leighton, M. A., Souza, P., & Straub, S. (2019). Social promotion in primary school: Effects on grade progression. *Brazilian Review of Econometrics, 39*(1), 1-33. https://doi.org/10.12660/bre.v39n12019.78513

* Lorence, J. (2014). Third-grade retention and reading achievement in Texas: A nine year panel study. *Social Science Research, 48*, 1-19. https://doi.org/10.1016/j.ssresearch.2014.05.001

* Lorence, J., & Dworkin, A. G. (2006). Elementary grade retention in Texas and reading achievement among racial groups: 1994–2002. *Review of Policy Research, 23*(5), 999-1033. https://doi.org/10.1111/j.1541-1338.2006.00247.x

* Mahjoub, M. B. (2017). The treatment effect of grade repetitions. *Education Economics, 25*(4), 418-432. https://doi.org/10.1080/09645292.2017.1283006

* Manacorda, M. (2012). The cost of grade retention. *Review of Economics and Statistics, 94*(2), 596-606. https://doi.org/10.1162/REST_a_00165

* Mariano, L. T., & Martorell, P. (2013). The academic effects of summer instruction and retention in New York City. *Educational Evaluation and Policy Analysis, 35*(1), 96-117. https://doi.org/10.3102/0162373712454327

* Martorell, P., & Mariano, L. T. (2017). The causal effects of grade retention on behavioral outcomes. *Journal of Research on Educational Effectiveness, 11*(2), 192-216. https://doi.org/10.1080/19345747.2017.1390024

* Mathys, C., Véronneau, M.-H., & Lecocq, A. (2017). Grade retention at the transition to secondary school: Using propensity score matching to identify consequences on psychosocial adjustment. *The Journal of Early Adolescence, 39*(1), 97–133. https://doi.org/10.1177/0272431617735651

* McCombs, J. S., Kirby, S. N., Mariano, L. T., & Education, R. (2009). *Ending social promotion without leaving children behind: The case of New York City*. https://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG894.pdf

* Mendez, L. M. R., Kim, E. S., Ferron, J., & Woods, B. (2015). Altering school progression through delayed entry or kindergarten retention: Propensity score analysis of long-term outcomes. *Journal of Educational Research, 108*(3), 186-203. https://doi.org/10.1080/00220671.2013.867474

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS med*, *6*(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Mons, N. (2007). *Les nouvelles politiques éducatives* [New educational policies]. Presses Universitaires de France.

* Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology, 104*(3), 603-621. https://doi.org/10.1037/a0027571

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Causal inference in educational and social science research*. Oxford University Press.

* Nunes, L. C., Balcão Reis, A., & Seabra, C. (2018). Is retention beneficial to low-achieving students? Evidence from Portugal. *Applied Economics, 50*(40), 4306-4317. https://doi.org/10.1080/00036846.2018.1444261

* Okurut, J. M. (2018). Automatic promotion and student dropout: Evidence from Uganda, using propensity score in difference in differences model. *Journal of Education and Learning, 7*(2), 191-209. https://doi.org/10.5539/jel.v7n2p191

* Ou, S. R., & Reynolds, A. J. (2010). Grade retention, postsecondary education and public aid receipt. *Educational Evaluation and Policy Analysis, 32*(1), 118-139. https://doi.org/10.3102/0162373709354334

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan: A web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

* Ozek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. *Education Finance and Policy, 10*(3), 350-377. https://doi.org/10.1162/EDFP_a_00166

* Park, S., Steiner, P. M., & Kaplan, D. (2018). Identification and sensitivity analysis for average causal mediation effects with time-varying treatments and mediators: Investigating the underlying mechanisms of kindergarten retention policy. *Psychometrika, 83*, 298–320. https://doi.org/10.1007/s11336-018-9606-0

* Pereira, M. C., & Reis, H. J. (2014). Grade retention during basic education in Portugal: Determinants and impact on student achievement. *Economic Bulletin, 16*(6), 61-83.

* Peterson, L. S., & Hughes, J. N. (2011). The differences between retained and promoted children in educational services received. *Psychology in the Schools, 48*(2), 156-165. https://doi.org/10.1002/pits.20534

Piaget, J., & Inhelder, B. (1962). *The psychology of the child*. Basic Books.

Reschly, A. L., & Christenson, S. L. (2013). Grade retention: Historical perspectives and new research. *Journal of School Psychology, 51*(3), 319-322. https://doi.org/10.1016/j.jsp.2013.05.002

* Roderick, M., & Nagaoka, J. (2005). Retention under Chicago's high-stakes testing program: Helpful, harmful, or harmless? *Educational Evaluation and Policy Analysis, 27*(4), 309-340. https://doi.org/10.3102/01623737027004309

* Schumann, N. (2012). *Grade retention and student achievement: What is the impact over time?* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.680.4776&rep=rep1&type=pdf

* Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics, 152*, 154-169. https://doi.org/10.3386/w21509

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. *Educational Research Review, 6*(1), 1-26. https://doi.org/10.1016/j.edurev.2010.07.002

* Troncin, T. (2005). *Le redoublement: Radiographie d'une décision à la recherche de sa légitimité.* [Doctoral dissertation, Université de Bourgogne]. https://tel.archives-ouvertes.fr/tel-00140531/document

* Uysal, S. D. (2010). *The effect of grade retention on school outcomes: An application of doubly robust estimation methods* [Paper presentation]. Jahrestagung des Vereins für Socialpolitik: Okonomie der Familie, Frankfurt, Germany. https://www.econstor.eu/bitstream/10419/37510/2/VfS_2010_pid_885.pdf

* Vandecandelaere, M., Schmitt, E., Vanlaar, G., De Fraine, B., & Van Damme, J. (2016). Effects of kindergarten retention for at-risk children's psychosocial development. *Educational Psychology, 36*(8), 1354-1389. https://doi.org/10.1080/01443410.2014.950194

* Vandecandelaere, M., Vanlaar, G., Goos, M., De Fraine, B., & Van Damme, J. (2011). *Zittenblijven in de derde kleuterklas: Effecten op de wiskundeprestatie (SSL rapport nr. 51) [Kindergarten retention: Effects on performance in mathematics (SSL report 51)]*. https://www.researchgate.net/publication/241207077_Zittenblijven_in_de_derde_kleuterklas_Effecten_op_de_wiskundeprestaties_SSL_rapport_nr_51

* Vandecandelaere, M., Vanlaar, G., Goos, M., De Fraine, B., & Van Damme, J. (2013). Effecten van zittenblijven in de derde kleuterklas op de wiskundegroei: Een propensityscore-stratificatie-analyse [Effects of kindergarten retention on mathematics development: A propensity score stratification analysis]. *Pedagogische Studiën, 90*(5), 58-73.

* Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). The effects of early grade retention: Effect modification by prior achievement and age. *Journal of School Psychology, 54*, 79-101. https://doi.org/10.1016/j.jsp.2015.10.004

* Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). Time-varying treatments in observational studies: Marginal structural models of the effects of early grade retention on math achievement. *Multivariate Behavioral Research, 51*(6), 843-864. https://doi.org/10.1080/00273171.2016.1155146

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. https://www.jstatsoft.org/v36/i03/

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

What Works Clearinghouse (2020). *What Works Clearinghouse standards handbook* (Version 4.1). https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf

* Winters, M. A., & Greene, J. P. (2012). The medium-run effects of Florida's test-based promotion policy. *Education Finance and Policy, 7*(3), 305-330. https://doi.org/10.1162/EDFP_a_00069

* Wu, W., West, S. G., & Hughes, J. N. (2008). Effect of retention in first grade on children's achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology, 100*(4), 727-740. https://doi.org/10.1037/a0013098

* Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology, 102*(1), 135-152. https://doi.org/10.1037/a0016664

Xia, N., & Kirby, S. N. (2009). *Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes.* http://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR678.pdf

4. STUDY 2: ONE STEP BACK OR ONE STEP FORWARD? EFFECTS OF GRADE RETENTION AND SCHOOL RETENTION COMPOSITION ON PORTUGUESE STUDENTS' PSYCHOSOCIAL OUTCOMES USING PISA 2018

This chapter is based on Pipa, J., & Peixoto, F. (2022). One Step Back or One Step Forward? Effects of Grade Retention and School Retention Composition on Portuguese Students' Psychosocial Outcomes Using PISA 2018. *Sustainability, 14*, 16573. https://doi.org/10.3390/su142416573. Results were also presented at the European Conference on Educational Research [ECER, (2021)].

Abstract

Grade retention is a common practice applied to academically struggling students within the Portuguese context. Studies investigating the psychological experiences of grade-retained students are still scarce. In addition, most studies tend to neglect the multilevel nature of the school context. This study examines the effects of grade retention in grades 1–9 on Portuguese students' psychosocial outcomes by the age of 15, using PISA 2018 data. Using a quasi-experimental design through full matching, we reduced the bias between 1362 retained and 4189 promoted students in relevant background variables. Results from the multilevel models showed that retained students, by the age of 15, present lower task orientation and school be-longing. In addition, we found that the high retention rates negatively relate to students' reading self-concept, task orientation, and school valuing and that school retention rates moderate the relationship between students' retention and the psychosocial variables considered. Overall, these findings suggest detrimental effects of grade retention and that grade retention also affects the promoted peers of retained students.

Keywords: grade retention; psychosocial outcomes; school retention composition; qua-si-experimental methods; multilevel modelling; PISA

Introduction

When students struggle to meet the academic expectancies or goals established for a specific grade level, one option is to allow them more time by retaining them in the same grade (Allen et al. 2009; Driessen 2020). This type of response to deal with students' heterogeneity, in terms of aca-demic competencies, is one of the most discussed educational strategies (Jimerson 2001). Furthermore, each country usually uses grade retention rates as a measure of educational quality and equity (Conselho Nacional de Educação [CNE], 2020; European Comission/EACEA, 2020; Organization for Economic Co-operation and Development [OECD], 2020).

Accompanying the long debate around grade retention effectiveness, most countries decreased their retention rates during the last decade. However, in countries such as Belgium, Luxembourg, Germany, Spain, and Portugal (OECD, 2020), grade retention is still a common strategy used to deal with students' low achievement (Dupriez et al., 2008; European Comission/EACEA, 2020; Goos, Schreier, et al., 2013; Marsh, 2016). This paper focuses on data from Portugal, where grade retention rates are among the highest in the Organization for Economic Co-operation and Development (OECD) countries. In these educational systems, grade retention is believed to bring several benefits to struggling students, such as giving them more time to develop and mature and to catch up on the learning materials (European Comission/EACEA, 2020; Goos, Van Damme, et al., 2013; Ikeda & Garcia, 2014). In addition, retaining students in grades is believed to improve teacher effectiveness by creating more homogenous classrooms, and the threat of being retained might boost students' motivation to work harder (Dupriez et al., 2008; Goos, Van Damme, et al., 2013; Hong & Raudenbush, 2005).

However, opponents of grade retention argue that grade retention harms students' motivation, confidence, and sense of self-worth because, as they will go over the same curriculum once more, they are deprived of intellectual challenges and meaningful learning (Hong & Yu, 2007). Moreover, retained students are detached from their peer group and will face a new classroom of younger students, leading to a decreased perceived competence and a sense of failure (Goos, Van Damme, et al., 2013). Finally, being retained brings extra economic and opportunity costs for the educational system, students, and families (Ehmke et al., 2010; Ikeda & Garcia, 2014).

Being retained could constitute a rather negative psychosocial experience for students (Anderson et al., 2005; Jimerson & Ferguson, 2007; Martin, 2011) and has been pointed out by children and adolescents as one of the most stressful life events (Anderson et al., 2005), affecting their motivation and self-confidence (Goos, Van Damme, et al., 2013; Hong & Yu, 2007; Martin, 2011; Peixoto et al., 2016; Peixoto et al., 2017). Nevertheless, despite its relevance for educational success and adjustment, studies considering the effects of grade

retention on students' psychosocial outcomes, such as self-concept, motivation, or engagement, have received less attention from the research community (Goos et al., 2021). In addition, as advised by other scholars, school context plays a crucial role in shaping students' self-beliefs and experiences (Marsh, 2016). In particular, school retention composition, i.e., the proportion of repeaters in a given school, was associated with academic and psychosocial outcomes, namely, students' peer relationships, self-concept, school belonging, and enrolment in post-secondary education (Demanet & Van Houtte, 2016; Van Canegem et al., 2021, 2022; Xiang & Chiu, 2022).

Therefore, this paper investigated whether Portuguese students with a history of grade retention differ from their promoted same-age peers in psychosocial outcomes. Moreover, we aimed to study whether these differences could also be exacerbated or not in schools with a more significant proportion of retained students. For this study, we relied on the available data from the Programme for International Study Assessment (PISA) 2018 regarding the Portuguese context.

Previous Research on the Effects of Grade Retention

Research on grade retention effectiveness has grown tremendously, especially since 2010 (Goos et al., 2021). Recent meta-analyses and systematic reviews estimated an overall null effect of grade retention (Allen et al., 2009; Goos et al., 2021; Valbuena et al., 2021). However, these studies found that the impact of grade retention is highly dependent on the context where it occurs (e.g., country, state, and educational system), the timing of grade retention, the timing of follow-up (e.g., short-term studies vs. long-term studies), and the variables observed (e.g., academic achievement vs. psychosocial or school career) (Goos et al., 2021; Valbuena et al., 2021).

The existing studies considering psychosocial outcomes are far less conclusive than those investigating academic achievement outcomes, suggesting positive, negative, or nonsignificant effects (Jimerson, 2001; Goos et al., 2021; Valbuena et al., 2021). These inconclusive findings could be attributed to the broader nature of the term 'psychosocial outcomes', covering different variables that could be differently affected by grade retention (Goos et al., 2021; Van Canegem et al., 2021). Additionally, short-term studies tend to present more positive results (Ehmke et al., 2010; Goos, Van Damme, et al., 2013; Klapproth et al., 2016; Peixoto et al., 2016; Van Canegem et al., 2021; Wu et al., 2010).

Effects on Students' Academic Self-Concept

Students' academic self-concept, defined as students' self-perception of competence in specific academic-related domains (e.g., reading self-concept), plays a significant role in school adjustment, achievement, and educational success (Marsh, 2016). To explain the

effects of grade retention on students' academic self-concept, researchers have often referred to the big-fish–little-pond effect (Marsh et al., 2008). This effect posits that students compare their own school-related accomplishments with those of their classmates, and this frame of reference act as the base for their self-concept development (Marsh, 2016; Marsh et al., 2017; Marsh et al., 2008)

Based on this framework, one could expect that retained students would develop more positive academic self-beliefs, at least during the retention year, because their frame of reference would comprise their younger grade mates with less academic experience (Marsh, 2016; Marsh et al., 2017). On the other hand, however, some scholars also claim that being retained jeopardises students' perception of competence because they may perceive that being retained constitutes a personal failure that makes them less competent and capable (Ikeda & Garcia, 2014; Jimerson, 2001).

Thus, the empirical evidence also shows mixed effects of grade retention. Longitudinal studies assessing grade retention in primary (Peixoto et al., 2016) and lower secondary education (Ehmke et al., 2010; Kretschmann et al., 2019; Lamote et al., 2014; Marsh et al., 2017; Peixoto et al., 2016) found either positive effects on math and academic self-concept (Ehmke et al., 2010; Lamote et al., 2014; Marsh et al., 2017), or adverse effects on language, math, and academic self-concept (Kretschmann et al., 2019; Peixoto et al., 2016), during the retention year. Conversely, more adverse long-run findings emerged from cross-sectional studies investigating effects in lower (Klapproth et al., 2016) and upper secondary education (Martin, 2011; Van Canegem et al., 2021). In contrast, longitudinal studies generally revealed nonsignificant effects on students' academic, language, and math self-concept during lower and upper secondary education (Hung & Yu, 2008; Klapproth et al., 2016; Kretschmann et al., 2019; Lamote et al., 2014; Marsh et al., 2017). Positive long-run effects were found in only one study using international PISA data (Marsh, 2016).

Effects on Students' Goal Orientations

Students' motivation has received much attention from researchers as it is recognised for its critical role in students' academic behaviour and performance (Meece et al., 2006). Students' goal orientations have become one of the largest research fields in motivation and are characterised by students' reasons or purposes for engaging in certain achievement behaviours (Kaplan & Maehr, 2007). The PISA in 2018 assessed two kinds of these reasons, or orientations, to engage in academic tasks: task orientations and self-enhancing ego orientations (Skaalvik, 1997). In task orientation, students engage in a school-related task (e.g., a reading task) to develop and acquire knowledge or master a new skill. On the other hand, when students pursue a self-enhancing ego orientation, the aim is to demonstrate competence and outperform others (Elliot & McGregor, 2001; Skaalvik, 1997).

Considering the effects of grade retention on motivational outcomes in general, one may expect, on the one hand, that grade retention acts as a 'boost' in students' motivation since they will finally experience success and will receive positive feedback from teachers (Goos, Van Damme, et al., 2013). Additionally, students may view the retention year as a second chance to master the learning content or even an opportunity for a fresh start (Ehmke et al., 2017; Kretschmann et al., 2019).

Despite its relevance within the educational context, to our knowledge, studies investigating the effects of grade retention on students' goal orientations are still scarce. The existing studies mainly suggest that students with a retention history demonstrate less adaptive motivational profiles in lower secondary education, even before being retained (Kretschmann et al., 2019, and during the retention year and beyond (Peixoto et al., 2016; Peixoto et al., 2017).

Effects on Students' Sense of School Belonging and Valuing

Apart from students' individual motivation, their social exchanges within the school context, particularly with their peers, teachers, and the broader school community, are essential agents in shaping their motivation at school (Goodenow & Grady, 1993). This social aspect of school motivation is commonly known as students' sense of school belonging (Finn, 1989; Goodenow & Grady, 1993; OECD, 2003; Slaten et al., 2016). Students' sense of school belonging can be understood as students' feelings of being accepted, respected, valued, and supported by their peers and the broader school community (Finn, 1989; Goodenow & Grady, 1993; OECD, 2003). In addition, these feelings of belonging are also associated with valuing school and school success, the two components of students' participation at school (Finn, 1989).

When students are retained in a grade, they lose their peer group. This experience of a broken relationship can trigger feelings of isolation and alienation from school (Demanet & Van Houtte, 2016; Goos, Van Damme, et al., 2013) due to failure to satisfy the need to establish and, especially, to maintain stable relationships with others (Slaten et al., 2016; Wentzel, 2002). In addition, grade retention is an explicit form of academic failure. This stigma of failing a grade and not being good enough academically can make establishing new and positive relationships in a new and unfamiliar peer group even harder (Demanet & Van Houtte, 2016). These feelings and experiences may thus lead to a greater sense of school disaffection and feelings of being an outsider from school and not connected with the school community (Demanet & Van Houtte, 2016; Goos, Van Damme, et al., 2013).

Empirical studies investigating the effects of grade retention on students' sense of belonging suggest that grade retention does not improve students' sense of school be-longing. Longitudinal studies showed predominantly adverse effects in both primary (Wu et al., 2010)

and lower secondary education (Im et al., 2013). Cross-sectional studies revealed mainly adverse effects of retention on school belonging (Van Canegem et al., 2021), particularly those using international PISA data (OECD, 2019; Van Canegem et al., 2022). To our knowledge, the effects of grade retention on students' school valuing as it is operationalised here are inexistent.

School Retention Composition

Empirical studies recognising the importance of school context in studying grade retention, although less common, have been growing during the last years, suggesting its crucial role in moderating the relationship between individual grade retention and academic and psychosocial outcomes (Demanet & Van Houtte, 2013, 2016; Hong & Raudenbush, 2005, 2006; Santos et al., 2022; Van Canegem, et al, 2021, 2022; Xiang & Chiu, 2022). In addition, Van Canegem et al. (2021) posited that the context where grade retention occurs might be crucial to clarify divergent findings from previous studies.

The number of retained students attending a particular school can impact students' psychosocial outcomes in two ways. First, a direct effect of school retention composition is expected through the so-called spillover effects of retained students on their non-retained peers (Van Canegem, et al, 2021). Spillover effects of grade retention have been less considered in grade retention research, despite constituting a big concern for families and educators (Goos et al., 2021). Retaining students in a particular classroom may negatively affect the classroom climate and instruction and, therefore, the learning of non-retained classmates (Gottfried, 2013; Xiang & Chiu, 2022).

Second, school retention composition may moderate the relationship between individual grade retention and students' psychosocial outcomes. Accordingly, retained students might present more negative behaviours and feelings of being stigmatised in schools with low rates of grade retention. These feelings might, thus, exacerbate the impact of grade retention on students' self-concept and sense of belonging, for example (Demanet & Van Houtte, 2013, 2016; Van Canegem et al., 2021, 2022).

The empirical studies that have addressed the impact of school retention composition on students' psychosocial factors showed that students from schools with many repeaters tend to be more likely to misbehave at school have a lower number of friends, and lower levels of academic self-concept and school belonging (Demanet & Van Houtte, 2013, 2016; Van Canegem et al., 2021, 2022). In addition, some studies showed more favourable results of grade retention in schools with higher retention rates (Demanet & Van Houtte, 2016; Van Canegem et al., 2021, 2022).

The Portuguese Context

The Portuguese school system, along with other southern European countries (e.g., France, Italy, and Spain), offers a common core curriculum for all students until 9th grade, and grade retention is the primary strategy applied to deal with academically struggling students and to promote homogeneity inside the classroom (Dupriez et al., 2008). In Portugal, 24% of 15-year-old students have reported being retained at least once during their school career, being largely above the OECD average of 11% (OECD, 2020). These rates may reflect the 'culture of retention' mentioned in several studies, meaning that educators believe that grade retention is beneficial for students over and above the recommendations of the international educational community and even national legislation (European Commission/EACEA, 2020; Goos, Schreier, et al., 2013). Currently, the national legislation states that grade retention in Portugal should only be an 'exceptional measure' when promoting the student to the next grade compromises the acquisition of new learnings (Decreto-Lei no. 55/2018). Thus, grade retention decision falls on schools and teachers who, except for transition years, have the autonomy to define in which specific circumstances grade retention will be applied.

Although it is a widespread practice in Portugal, studies using Portuguese data are still scarce, especially in using adequate methodologies and considering variables beyond student achievement (Nunes et al., 2018; Pereira & Reis, 2014). To our knowledge, only a limited number of studies have focused on the effects of grade retention on students' psychosocial outcomes, such as students' self-concept and motivation. These studies showed predominantly adverse effects (Peixoto et al., 2016, 2017; Santos et al., 2022).

The Present Study

In summary, the following considerations guided this study: (a) the findings that the characteristics of each school system moderate the effects of grade retention; (b) the limited number of empirical studies evaluating grade retention effects within the Portuguese context (and to some extent considering other countries with a similar educational system); (c) the mixed empirical evidence of the effects of grade retention on students' psychosocial variables driven by the differential nature of each construct; and (d) the importance of school retention composition in clarifying the relationship between grade retention and academic outcomes.

Hence, we aimed to investigate the effects of grade retention on a group of psychosocial variables. Specifically, using a large-scale assessment and applying a same-age comparison approach, we examined (1) whether students who had experienced grade retention at least once between grade 1 and grade 9 differed from their same-age pro-moted peers in reading self-concept, goal orientations, and students' sense of school belonging and valuing; and (2) whether the nature of these effects can differ according to the proportion of retained students attending a school, i.e., school retention composition.

Method

Data and Participants

Starting in 2000, PISA became one of the largest and most prominent large-scale assessment studies in education. Every three years, PISA assesses reading, math, and science competencies acquired by 15-year-old students. In addition, PISA gathers in-formation regarding student and family background information and various psycho-social variables, including academic self-concept, motivation, and school engagement.

This study used data from 5932 Portuguese students who participated in PISA 2018, attending 276 schools. Since this study focused on grade retention effects, we excluded students who had missing information (5%, n = 308) in the grade retention variable from the data. Thus, further analyses were based on 1362 retained (24%) and 4262 promoted students (M_{age} = 15.73, SD_{age} = 0.29, 50% boys), attending between 7th and 11th grade. PISA uses a two-stage process regarding the sampling procedure to obtain a representative sample of students from each country and economy. First, schools are randomly selected from a complete list of schools containing the student population of interest. Second, thirty-five 15-year-old students from each school are randomly selected to fulfil the questionnaires (OECD, 2019a).

Measures

In this study, we focused on PISA data considering students' reading self-concept, goal orientations, and school belonging and valuing. In addition, we used information regarding students' social background and school characteristics retrieved from students' and school principals' questionnaires. All continuous measures were standardised to have a mean of 0 and a standard deviation of 1, and categorial measures were coded as dummy variables to facilitate interpretation. Concerning the validity and reliability of the measures used, internal consistency ranged from α = 0.74 to α = 0.88 (see Table 1), revealing acceptable levels of reliability, as was also referred to in the PISA report (OECD, 2019b).

Grade Retention

In the PISA questionnaire, students were asked whether they had ever repeated a grade in ISCED I, II, or III. This study considered grade retention responses regarding ISCED I or II. In these variables, 0 means that a student never repeated a grade during ISCED I or II, whereas 1 means that a student has repeated it at least once. In addition, school retention composition was derived from responses to grade retention in ISCED II and III variables since many schools offer both these two levels of education and were operationalised as the percentage of retained students in each school.

Reading Self-Concept

To assess academic self-concept in reading, students were asked on a 4-point scale (strongly disagree to strongly agree) whether they perceive themselves as good readers, whether they are able to understand complex texts, and whether they read fluently (OECD, 2019c).

Goal Orientations

Students' task orientations, or learning goals as mentioned by PISA (OECD, 2019c, p. 215), were measured on a 5-point scale (not at all true of me to extremely true of me), asking to what extent they have the goal of learning and master class-related materials. Likewise, to assess self-enhancing orientations or attitudes toward competition, as referred to by PISA [OECD, 2019c, p. 215), students answered three items on a 4-point scale (strongly disagree to strongly agree), asking whether they enjoy working in competitive environments and whether they have the goal of outperforming others.

School Belonging and Valuing

To assess their sense of school belonging, students answered six statements, such as "I feel like an outsider (or left out of things) at school" and "I feel like I belong at school" (OECD, 2019b, p.130), on a 4-point scale (strongly disagree to strongly agree). Students also completed three items to assess how much they value school. In these statements, students were asked whether they agree (strongly disagree to strongly agree) that trying hard at school would help them obtain a good job or help them be accepted into a good college (OECD, 2019b).

Students' Social Background and Competencies

We considered several individual background variables retrieved from the students' questionnaire and the measurement of competencies that were related to grade retention (Agasisti & Cordero, 2017; Choi et al., 2018; Ehmke et al., 2010; Klapproth & Schaltz, 2015; Marsh, 2016; Nunes et al., 2018; Pires et al., 2021). Specifically, we retrieved information considering students' age, gender, immigrant background, language spoken at home, index of economic, social, and cultural status (derived from parents' highest level of education, parents' highest occupational status, and home possessions), home educational resources (i.e., household possessions and the number of books at home), and parents' emotional support (self-report measure where students where asked whether they feel supported by their parents; OECD, 2019c) . In addition, we considered students' PISA scores in reading, math, and science.

Table 1

Means and Standard Deviations for Retained and Promoted Students, Internal Consistency, and Correlations Between the Outcome Variables

	Repeaters		Non-repeaters		Internal consistency					
	M	*SD*	*M*	*SD*	α	1	2	3	4	5
1.Reading self-concept	-0.50	0.83	-0.16	0.83	.74	-				
2.Task orientation	-0.33	0.99	0.11	0.93	.83	.25***	-			
3.Self-enhancing orientation	-0.04	0.96	-0.06	1.00	.87	.06***	.08***	-		
4.School belonging	-0.02	0.97	0.17	0.97	.80	.13***	.17***	.10***	-	
5.School utility value	0.09	0.97	0.50	0.85	.88	.14***	.32***	.01	.19***	-

***$p < .001$

School Context

We also integrated some school-related components as covariates considered in the literature related to grade retention (Agasisti & Cordero, 2017; Choi et al., 2018; Klapproth & Schaltz, 2015; Pires et al., 2021; Xiang & Chiu, 2022). Specifically, we retrieved information regarding school type (public vs. private; OECD, 2019c). The proportion of participating students attending public schools does not differ from the true proportion of Portuguese students attending such schools (88% in both cases; CNE, 2020). In addition, school composition in terms of the index of economic, social, and cultural status and immigrant background, obtained by aggregating students' responses, was also considered.

Data Analysis

Handling Missing Values

In most observational studies, participants often leave one or more questions unanswered. Researchers have been encouraged to deal with incomplete datasets in recent years by applying imputation methods. In this study, we applied multiple imputations by chained equations using the *MICE* package in R (van Buuren & Groothuis-Oudshoorn, 2011), generating five completed datasets and allowing ten iterations. The proportion of missing values in the variables used in the analyses ranged from 0.1% to 10%. Subsequent analyses were conducted in each imputed dataset and then aggregated (Rubin, 1987).

Group Comparison Strategy

Applying an experimental methodology to estimate the causal effects of grade retention is not theoretically or ethically attainable, as students cannot be randomly retained or promoted. Nevertheless, the literature on grade retention effects draws attention to the importance of establishing comparability between the intervention (i.e., retained students) and comparison (i.e., promoted students) groups (Allen et al., 2009; Goos et al., 2021). Thus, propensity score matching methods are often applied to reduce selection bias by achieving a balance between treatment and comparison groups regarding background characteristics related to both the treatment and the outcome (Austin, 2011; Cordero et al., 2018). These methods have also been widely used in international studies, such as those using the PISA data (Cordero et al., 2018). We opted to apply the full matching technique among the different matching methods, using the MatchIt package in R (Ho et al., 2011). The full matching technique is considered a more sophisticated and flexible matching method and has the advantage of not discarding any observation (as does, for example, the one-to-one nearest neighbour technique). Full matching forms a series of matched sets (subclasses) containing at least one treated and one comparison subject (Stuart, 2010). After creating these matched sets, each comparison individual receives a weight proportional to the number of treatment

individuals present in each set (Ho et al., 2011). In subsequent outcome analyses, these weights are introduced in weighted regression models (Ho et al., 2011). Full matching techniques are recognised to be efficient in maximising the similarities between treatment and comparison individuals in each matched set (Stuart, 2010) and have been success-fully used in grade retention effects research (Kretschmann et al., 2019). The final step of preparing comparison groups for the outcome analysis was assessing the balance between treated and com-parison subjects across the covariates used (Stuart, 2010). This assessment was performed considering (Ho et al., 2011; Stuart, 2010): (a) standardised mean differences below 0.25; (b) variance ratios between 0 and 2; and (c) graphical inspection. In this study, we include in the propensity score matching a series of background characteristics and interaction terms, described in sections 2.2.5. The selection of these variables was theoretically based, as mentioned previously.

Outcome Analysis

The effects of grade retention on psychosocial components were estimated using multilevel models since students are nested within schools (Marsh, 2016). Thus, a series of hierarchical linear models were computed for each outcome, considering four stages (Model 0 to Model 3). Model 0 or the 'null model' was estimated with no predictors to examine the amount of variance in the outcomes that is explained by either student or school levels. Model 1 was estimated by entering students' background variables, reading, math, and science scores, school context variables, and grade retention variables. In Model 2, school retention composition was added to the previous variables. Model 3 assessed the inter-action between grade retention and school retention composition. In estimating these models, we followed Stuart's (2010) recommendations of combining matching and regression methods by including in the regression models the predictors and controlling variables previously considered in the matching procedures. We used weighted regression models, as already mentioned, considering the weights obtained after matching. Finally, in every model, we additionally checked for multicollinearity using the variance inflation factor (VIF). Using a threshold of 10 indicating a strong correlation between the independent variables, our results supported its independence, with VIF values ranging between 1.02 and 4.84.

Results

Covariate Balance across Retained and Promoted Students

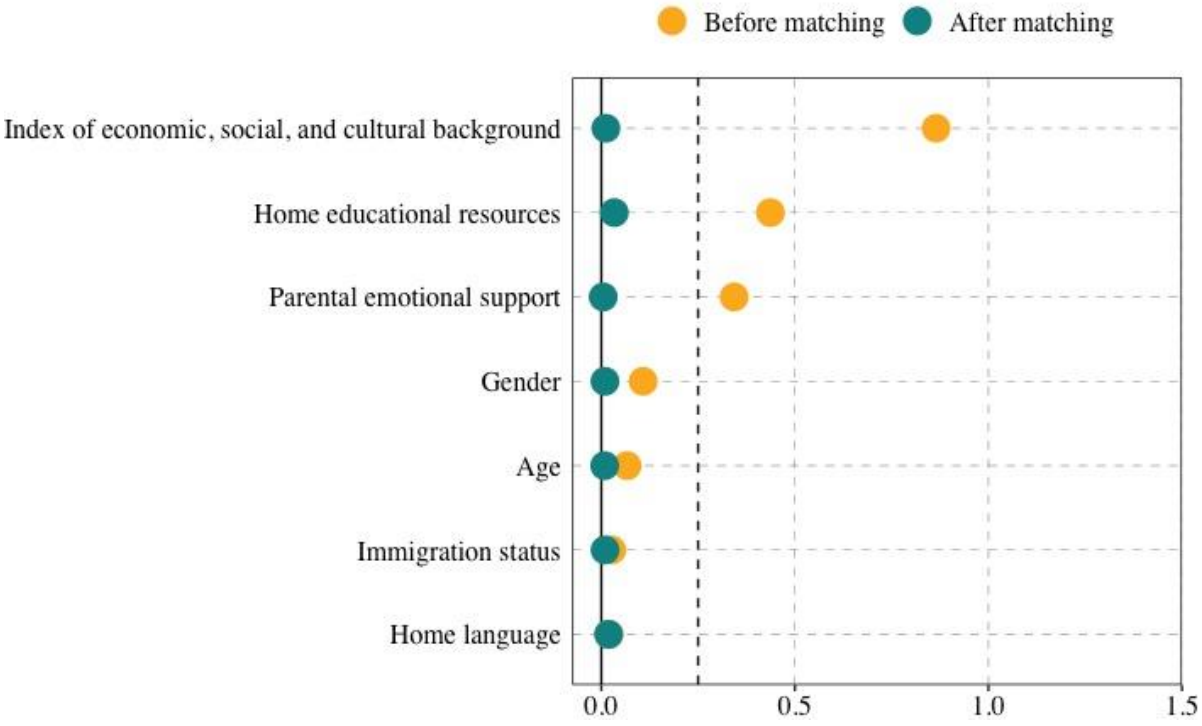
Before conducting the outcome analysis, we assessed whether the selection bias on background variables was reduced through full matching. It is first worth mentioning that we observed extreme weights in some comparison individuals, suggesting difficulties in finding suitable matching individuals for these observations (Stuart, 2010). Hence, we discarded the

observations with extreme weights above the 99th percentile (*n* = 73). Further analyses were based on 1362 retained students and 4189 promoted students.

Results from the full matching revealed adequate balance. As shown in Figure 1, the absolute standardised mean differences between retained and promoted students in the background variables decreased considerably after matching, with all standardised mean differences below 0.25. In critical, highly related to retention, background variables, such as students' ESCS, the standardised mean differences were reduced to 0.01 standard deviations.

Figure 1

Absolute Standardised Mean Differences Between Retained (n = 1362) and Promoted Students (n = 4189) Before and After Full Matching



Effects of Grade Retention

This section presents the effects of grade retention on the outcome variables assessed. Table 1 shows descriptive statistics in the outcome variables for the retained and promoted groups and correlations between outcome variables. To improve readability, in Table 2, we present only the results from Model 3, i.e., the model testing both single effects of grade retention and school retention composition and interaction effects between these variables. In the following sections, we briefly describe the findings from the remaining models. Detailed information regarding coefficients from Model 0 to Model 2 for each outcome variable can be found in Appendix B.

Table 2

Multilevel Unstandardized Parameter Estimates From Model 3 for the Outcomes Assessed

	Reading self-concept	Task orientation	Self-enhancing orientation	School belonging	School utility value
Intercept	-0.10 (0.08)	0.12 (0.10)	-0.05 (0.10)	0.45 (0.10)***	0.49 (0.10)***
Student level					
Retention	0.01 (0.03)	-0.16 (0.03)***	-0.07 (0.03)	-0.12 (0.03)***	-0.08 (0.03)*
Age	0.01 (0.01)	0.02 (0.01)	0.01 (0.01)	-0.03 (0.01)**	-0.06 (0.01)***
Male	-0.06 (0.02)*	-0.21 (0.03)***	0.42 (0.03)***	0.20 (0.03)***	-0.16 (0.03)***
ESCS[a]	0.08 (0.01)***	0.02 (0.02)	0.02 (0.02)	0.06 (0.02)***	0.02 (0.02)
Immigration status	0.03 (0.04)	0.20 (0.05)***	-0.14 (0.06)*	-0.18 (0.05)**	0.09 (0.05)
Home language	-0.26 (0.06)***	0.05 (0.07)	-0.17 (0.07)*	0.01 (0.07)	0.06 (0.07)
HEDRES[b]	0.10 (0.01)***	0.16 (0.01)***	0.04 (0.02)*	0.09 (0.01)***	0.11 (0.01)***
EMOSUPP[c]	0.07 (0.01)***	0.21 (0.01)***	0.11 (0.01)***	0.23 (0.01)***	0.22 (0.01)***
Reading score	0.37 (0.03)***	-0.10 (0.04)**	-0.14 (0.04)***	0.10 (0.03)**	0.24 (0.03)***
Math score	-0.31 (0.03)***	-0.07 (0.04)	-0.21 (0.04)***	-0.30 (0.03)***	0.03 (0.03)
Science score	0.12 (0.03)***	0.20 (0.04)***	0.34 (0.04)***	0.29 (0.04)***	-0.10 (0.04)**
School level					
Retention composition	-0.07 (0.03)*	-0.13 (0.04)***	-0.01 (0.03)	-0.05 (0.03)	-0.12 (0.03)***
Public school	0.11 (0.07)	-0.07 (0.08)	-0.01 (0.08)	-0.32 (0.08)***	-0.03 (0.07)
School ESCS[a] composition	-0.04 (0.03)	-0.04 (0.03)	-0.01 (0.03)	0.02 (0.03)	0.01 (0.03)
School immigrant composition	-0.01 (0.02)	0.01 (0.02)	-0.01 (0.02)	0.02 (0.02)	0.01 (0.02)

Retention X Retention composition	0.08 (0.02)***	0.13 (0.03)***	0.03 (0.03)	0.09 (0.03)**	0.14 (0.03)***
Between school variance	0.08 (0.28)	0.10 (0.32)	0.10 (0.31)	0.10 (0.32)	0.09 (0.30)
Within school variance	0.31 (0.56)	0.47 (0.68)	0.51 (0.72)	0.43 (0.65)	0.43 (0.66)

Note. Standard errors are in parenthesis. [a]Index of economic, social, and cultural background; [b]Home educational resources; [c]Parental emotional support.

*$p < .05$, ** $p < .01$, ***$p < .001$

Effects of Grade Retention on Reading Self-Concept

The intraclass correlation coefficient (ICC) obtained from Model 0 (these coefficients are only described in the text; calculations are possible from the tables in Appendix A) suggests substantial variance between schools in reading self-concept (ICC = 0.20). Thus, these results support the multilevel analysis. Considering the main effect of grade retention, the results were consistent across the three models; the relationship between grade retention and reading self-concept was nonsignificant. Conversely, Model 3 revealed a negative relationship between school retention composition and reading self-concept ($b = -0.07$, $p < 0.05$), suggesting that students in schools with a larger pro-portion of repeaters tend to feel less competent in reading.

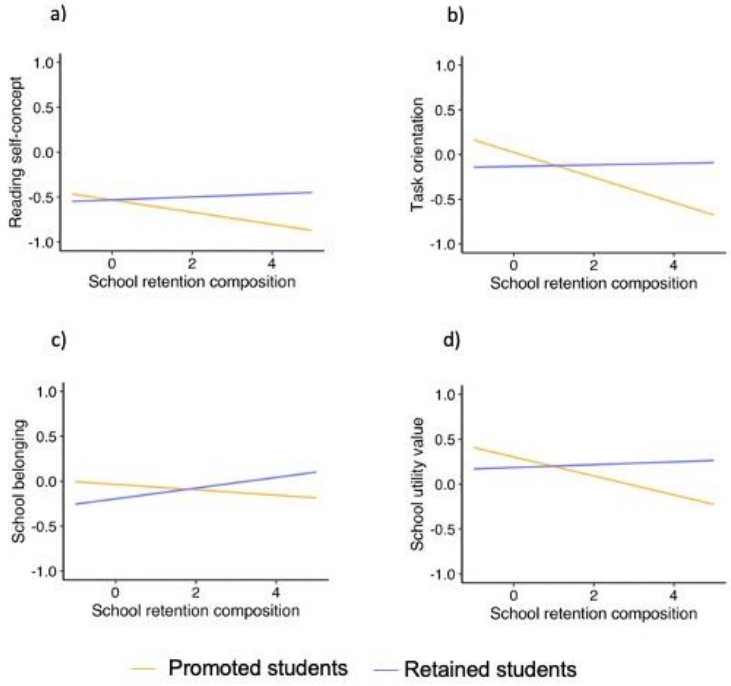
In addition, the results showed a combined effect of grade retention and school retention composition ($b = 0.08$, $p < 0.001$) on this outcome, revealing that the higher proportion of repeaters in school attenuates the effects of grade retention on students' reading self-concept. Figure 2a illustrates the nature of this interaction effect, showing that in schools with a high number of repeaters, repeaters tend to present a higher perception of reading competence than promoted students.

Effects of Grade Retention on Goal Orientations

Considering students' task orientation and self-enhancing orientation, the ICCs from Model 0 revealed that 19% and 16% of the variance in students' goal orientations is attributable to schools. The results from all models showed that retention is related to lower levels of task orientation ($b = -0.12$, $p < 0.001$; $b = -0.11$, $p < 0.001$; $b = -0.16$, $p < 0.001$; Models 1 to 3, respectively). Considering the effects of school retention composition, we found an effect of this variable only in Model 3 ($b = -0.13$, $p < 0.001$) when the interaction term was entered. The moderating effect of school retention composition in the relationship between individual retention and task orientation ($b = 0.13$, $p < 0.001$) showed that, on the one hand, grade retention affects students' task orientation less when they attend a school with a higher number of retained students. On the other hand, an inspection of this interaction considering retained and non-retained students showed that for retained students, being in a school with high retention rates does not affect or even slightly improve their task orientation and, for promoted students, being part of such a school negatively affects their task orientation (see Figure 2b).

Figure 2

Graphical Representation of the Interaction Effect Between Grade Retention and School Retention Composition that Resulted From Multilevel Hierarchical Models (Model 3).



Note. (a) Reading self-concept; (b) Task orientation; (c) School belonging; (d) School utility Value

Effects of Grade Retention on School Belonging and Valuing

For school belonging and utility value, the ICCs again showed substantial variability between schools in these components (ICC = 0.23 and ICC = 0.18, respectively). The results for school belonging showed in both models that grade retention is related to lower feelings of school belonging (b = −0.09, p = 0.004, for Models 1 and 2; b = −0.12, p < 0.001, Model 3). In addition, consistent with the previous results, individual retention was found to be less detrimental to students' sense of belonging when schools present many repeaters (b = 0.09, p = 0.001). Moreover, promoted students attending a school with a high rate of retained students present lower sense of school belonging (see Figure 2c). For students' sense of school utility value, this component was only affected by grade retention and school retention composition when the interaction term was considered (in-dividual retention: b = −0.03, p = 0.324, Model 1; b = −0.03, p = 0.415, Model 2; b = −0.08, p = 0.011, Model 3; retention composition: b = −0.12, p < 0.001). Again, school retention composition was found to moderate the effects of individual retention (b = 0.14, p < 0.001), showing that grade retention rates at the school level tend to attenuate the negative relationship between grade retention and school utility value. Moreover, in line with previous findings, promoted students attending a school

with a high rate of retained students present lower sense of school utility value (see Figure 2d).

Discussion

In Portugal, grade retention is still a prevalent practice applied to academically struggling students, irrespective of being considered an 'exceptional measure'. Therefore, based on data from PISA 2018, we aimed to explore the effects of grade retention on students' reading self-concept, goal orientations, and school belonging and valuing by employing methods that consider several background variables related to grade retention. The use of full matching (Stuart, 2010) and the available data on several individual and contextual variables allowed us to reduce the differences between promoted and retained students and to estimate more rigorous effects of grade retention by disentangling the impact of potential confounders, as previously recommended (Allen et al., 2009; Goos et al., 2021; Valbuena et al., 2021).

Effects of Grade Retention and School Retention Composition on Psychosocial Outcomes

In this study, we found that retained students, by the age of 15, are less oriented to master academic-related tasks and have a lower sense of school belonging and valuing. These results are in line with previous studies, suggesting a detrimental effect of grade retention on motivational variables and school belonging (Klapproth et al., 2016; Martin, 2011; OECD, 2019a; Peixoto et al., 2016; Peixoto et al., 2017; Van Canegem et al., 2021, 2022). Although the cross-sectional nature of this study limits our potential to make causal statements, our findings, together with the previous longitudinal and retrospective studies, suggest that retained students could engage in a negative cascade and that grade retention could leave an irreversible mark on students' motivation and engagement (Anderson et al., 2005; Jimerson & Ferguson, 2007). We additionally have explored the effects of repeating more than once by conducting separate analyses for students retained only once (n = 778) and students retained at least twice (n = 584). The analyses considering students retained only once revealed similar findings for all the outcomes. On the other hand, repeating more than once had positive effects on academic self-concept, negligible effects on task orientation and negative effects on students' self-enhancing orientation. The effects of school retention composition remained unaltered. Detailed results of these analyses are available on request from the first author.

These results are supported by the social goals framework (e.g., Wentzel, 2002), stressing that students are most likely to engage in a context that provides opportunities to meet the social goal of establishing personal relationships with teachers and peers. Thus, in the case of retained students, by losing their reference peer group and friends, they could experience feelings of not being accepted, respected, or valued within their school community.

In addition, students' motivation to learn is also affected by the fulfilment of social goals (Wentzel, 2002). In this study, this is visible in the low levels of task orientation and school utility value presented by retained students, suggesting that these students are less oriented to develop learning task-related skills and gradually devaluate school learning.

Regarding students' reading self-concept, our results seem to unravel a more com-plex picture than individual grade retention effects alone. For this variable, we have found a more substantial effect of school retention composition over and above the impact of individual retention on students' perception of competence. These findings could also be interpreted in light of the big-fish–little-pond effect (Marsh et al., 2008), which posits the school context's prevalence in shaping students' self-concept.

Indeed, one of the most notable findings of this study is the impact that school retention composition exerts on students' self-concept, motivation, and engagement. Overall, the larger the share of retained students in a school, the lower the self-perception of competence in reading, task orientation, and sense of school valuing. In addition, the interaction effect between individual retention and school retention composition suggests that the retention rates do not affect students equally. In the case of our study, non-retained students were those that presented more negative outcomes by attending a school with a large proportion of repeaters. Thus, the adverse effects of grade retention also have significant implications for classmates of retained students and the broader school community, supporting previous findings (Bietenbeck, 2014; Demanet & Van Houtte, 2013, 2016; Gottfried, 2013; Van Canegem et al., 2021).

These findings seem to reflect one of the major concerns of educators and parents—that sharing a classroom with repeaters by disrupting classroom instruction negatively affects the academic outcomes of their non-retained peers (Lavy et al., 2012). Thus, interventions, such as ability grouping, may sound attractive to ensure non-retained students' academic success. However, as previous studies showed, educational systems that preconise tracking and ability grouping in classes present the most detrimental effects of grade retention on students' development (Goos et al., 2021; OECD, 2020). Moreover, retention rates and individual retention are related to students' background and achievement, meaning that schools with a large proportion of repeaters are most often attended by socially disadvantaged and low-achieving students (European Commistion/EACEA, 2020; Lavy et al., 2012; Xiang & Chiu, 2022), and these are simultaneously the characteristics that put students in a more vulnerable position to being retained (Choi et al., 2018; Klapproth et al., 2016; Nunes et al., 2018). Therefore, the most vulnerable students are, simultaneously, more likely to be retained and to share a class/school with many repeaters (Gottfried, 2013), resulting in greater inequalities among these students (Xiang & Chiu, 2022). Here, tracking and ability grouping will only exacerbate this effect and, consequently, student disparities. Finally, at a macro level,

the PISA 2018 data revealed that countries and economies presenting higher grade retention rates generally showed lower levels of reading performance and lower levels of equity in education (OECD, 2020).

Limitations and Future Directions

The contributions of the findings from this study are not without their limitations. First, we must mention the cross-sectional and retrospective nature of the presented data that prevents us from investigating developmental trajectories regarding students' academic self-concept, goal orientations, sense of belonging and valuing, and establishing causal relationships. In the case of grade retention research considering Portuguese data, longitudinal studies regarding grade retention effectiveness would be very important given the limited number of studies within this context and the high retention rates. Further investigations assessing Portuguese retained students' developmental trajectories of academic and psychosocial aspects are needed to clarify the effectiveness of this practice.

Second, and linked with the previous limitation, it was impossible to disentangle which grade students were retained considering the data used in this study. Thus, we could not estimate the potential long-term effects of grade retention more precisely. Considering the specific grade when students were retained is important not only from a developmental perspective but also due to findings from previous studies, where grade retention effects differed according to the grade where they were retained (e.g., Giano et al., 2022).

Third, concerning the moderating effect of school retention composition, our operationalisation of this variable is limited because only a small number of students from each school participated in the PISA assessment, and non-identification of the participating schools prevents us from obtaining school retention rates from official records. Based on this, one should interpret our findings regarding school retention composition with caution. Given its notable contribution, we encourage researchers to consider this vital variable in future studies assessing grade retention effects.

Fourth, and finally, we assumed the broad definition of grade retention, considering it as a single and universal treatment. Additional interventions coupled with grade retention during the retention year and beyond, such as additional support and educational services provided to retained students, were not considered. Past research showed that positive effects could emerge when retention is coupled with other treatment sources (Goos et al., 2021). Moreover, we did not investigate the potential moderating effects of students' background characteristics. Some researchers suggest that students from certain subgroups or presenting specific features might benefit from grade retention (Choi et al., 2018; Goos et al., 2021). Although these aspects were beyond the scope of this study, we encourage researchers to collect information about 'what happens' during the retention year and test the effects of grade

retention according to some students' characteristics in future studies. These two considerations will be very informative either to research or to practice.

Conclusions

The findings of this study suggest that grade retention is not an effective practice. Specifically, our results indicate that grade retention is related to lower levels of motivation and engagement, two valuable conditions for school success. Moreover, it was reported that grade retention affects retained students and could also be detrimental to their peers attending the same school.

Based on our findings, we cannot support the use of grade retention as an effective intervention for struggling students. Instead, we first recommend early identification of at-risk students, monitoring their academic and psychosocial development, and providing additional support to avoid grade retention. Valbuena et al. (2021) listed numerous evidence-based and cost-effective policies, practices, and interventions that are alternatives to retention, such as tutoring, summer schools, and multi-age grouping. Likewise, students' academic competencies and psychosocial development should also be considered when deciding to retain a student, not only school marks. Furthermore, in the case of retaining a student, both these competencies should be monitored and supported equally, ensuring that the 'second chance' given to the students will not be harmful to their academic and psychosocial growth. To the broader school community, we recommend monitoring and reducing school retention rates since they affect the whole student community. At last, we advise educators and policymakers to continuously consider the psychosocial components of learning when debating the effectiveness of grade retention and its related policies and norms.

References

Agasisti, T., & Cordero, J. M. (2017). The determinants of repetition rates in Europe: Early skills or subsequent parents' help? *Journal of Policy Modeling, 39*, 129-146. https://doi.org/10.1016/j.jpolmod.2016.07.002.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis, 31*, 480–499. https://doi.org/10.3102/0162373709352239.

Anderman, E., & Wolters, C. (2006). Goals, values and affect: Influences on students motivation. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 369-389). Mahwah: Lawrence Erlbaum Associates.

Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005). Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology, 21*, 1–20. https://doi.org/10.1300/J370v21n01_01.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*, 399–424. https://doi.org/10.1080/0027 3171.2011.568786.

Bietenbeck, J. (2014). Learning from adversity? Short- and long-term spillover effects from grade retention in kindergarten. Retrieved from https://www.researchgate.net/publication/268447733_Learning_from_Adversity_Short-and_Long-Term_Spillover_Effects_from_Grade_Retention_in_Kindergarten

Choi, A., Gil, M., Mediavilla, M., & Valbuena, J. (2018). Predictors and effects of grade repetition in Spain. *Revista de Economia Mundial, 48,* 21-42. https://doi.org/10.33776/rem.v0i48.3882

Conselho Nacional de Educação (2020). *Estado da educação 2019* [State of Education 2019]. Conselho Nacional de Educação. https://www.cnedu.pt/pt/publicacoes/estado-da-educacao/1615-estado-da-educacao-2020

Cordero, J. M., Cristóbal, V., & Santín, D. (2018). Causal inference on education policies: A survey of empirical studies using PISA, TIMSS and PIRLS. *Journal of Economic Surveys, 32*, 878-915. https://doi.org/10.1111/joes.12217.

Decreto-Lei No. 55/2018. Diário da República – Série I – No. 129 de 6 de Julho, pp. 2928-2943. https://data.dre.pt/eli/dec-lei/55/2018/07/06/p/dre/pt/html

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement, 24*, 417–434. https://doi.org/10.1080/09243453.2012.727834

Demanet, J., & Van Houtte, M. (2016). Are flunkers social outcasts? A multilevel study of

grade retention effects on same-grade friendships. *American Educational Research Journal, 53*, 745– 780. https://doi.org/10.3102/0002831216646867

Driessen, G. (2020). *Grade retention, grade repetition, holding back a grade.* Retrieved from https://encyclopedia.pub/item/revision/ bcdab43e3cb9673d0d197cef8a9b4c18.

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review, 52*, 245–273. https://doi.org/10.1086/528764.

Elliot, A. J., & McGregor, H. A. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501-519. https://doi.org/10.1037/0022-3514.80.3.501.

Ehmke, T., Drechsel, B., & Carstensen, C. H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. *Studies in Educational Evaluation, 36*, 27–35. https://doi.org/10.1016/j.stueduc.2010.10.003.

Ehmke, T., Salzer, C., Pietsch, M., Drechsel, B., & Muller, K. (2017). Competence development in the school year after PISA 2012: Effects of grade retention. *Zeitschrift fur Erziehungswissenschaft, 20*, 99–124. https://doi.org/10.1007/s11618-017-0752-4.

European Commission/EACEA (2020). Equity in school education in Europe. Structures, policies and student performance. Eurydice. https://data.europa.eu/doi/10.2797/658266

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*(2), 117–142. https://doi.org/10.3102/00346543059002117.

Giano, Z., Williams, A. L., Becnel, J. N. (2022). Grade retention and school dropout: Comparing specific grade levels across childhood and early adolescence. *Journal of Early Adolescence, 42*, 33–57. https://doi.org/10.1177/02724316211010332.

Goodenow, C., & Grady, K. E. (1993). The relationship of school belonging and friends' values to academic motivation among urban adolescent students. *Journal of Experimental Education, 62*, 60–71. https://doi.org/10.1080/00220973.1993.9943831.

Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review, 34*. 100401. https://doi.org/10.1016/j.edurev.2021.100401

Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review, 57*, 54-84. https://doi.org/10.1086/667655

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade

retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*, 323–347. https://doi.org/10.1016/j.jsp.2013.03.002.

Gottfried, M. A. (2013). The spillover effects of grade-retained classmates: Evidence from urban elementary schools. *American Journal of Education 119*, 1– 64. https://doi.org/10.1086/669851.

Ho, D., Imai, K., King, G., & Stuart, E. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software, 42*, 1-28. https://doi.org/10.18637/jss.v042.i08.

Hong, G. L., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*, 205–224. https://doi.org/10.3102/01623737027003205.

Hong, G. L., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*, 901–910. https://doi.org/10.1198/016214506000000447.

Hong, G. L., & Yu, B. (2007). Early-grade retention and children's reading and math learning in elementary years. *Educational Evaluation and Policy Analysis, 29*, 239–261. https://doi.org/10.3102/0162373707309073.

Hong, G. L., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44*, 407–421. https://doi.org/10.1037/0012-1649.44.2.407.

Im, M. H., Hughes, J. N., Kwok, O. M., Puckett, S., & Cerda, C. A. (2013). Effect of retention in elementary grades on transition to middle school. *Journal of School Psychology, 51*, 349–365. httpw://doi.org/10.1016/ j.jsp.2013.01.004.

Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies, 2013*, 1–47. https://doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx.

Jimerson, S. R. (2001). Meta-analysis of grade retention research: Implications for practice in the 21st century. *School Psychology Review*, 30, 420-437. https://doi.org/10.1080/02796015.2001.12086124

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22*, 314-339. https://doi.org/10.1037/1045-3830.22.3.314

Kaplan, A., & Maehr, M. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review, 19*, 141-184.

https://doi.org/10.1007/s10648-006-9012-5.

Klapproth, F., & Schaltz, P. (2015). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. *European Journal of Psychology of Education, 30*, 119–136. https://doi.org/10.1007/s10212-014-0232-7.

Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences, 50*, 182–194. https://doi.org/10.1016/j.lindif.2016.08.014.

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology, 111*, 1432–1446. https://doi.org/10.1037/edu0000353.

Lamote, C., Pinxten, M., Van Den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies, 40*, 496–514. http://doi .org/10.1080/03055698.2014.936828.

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *Economic Journal, 122*, 208-237. https://doi.org/10.1111/j.1468-0297.2011.02463.x.

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology, 108*, 256–273. https://doi.org/10.1037/edu0000059.

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109*, 425-438. https://doi.org/10.1037/edu0000144.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*, 319–350 (2008). https://doi.org/10.1007/s10648-008-9075-6.

Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal, 37*, 739–763. https://doi.org/10.1080/01411926.2010.490874.

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student

motivation and academic achievement. *Annal Review of Psychology, 57*, 48–503. https://doi.org/10.1146/annurev.psych.56.091103.070258

Nunes, L. C., Balcão Reis, A., & Seabra, C. (2018). Is retention beneficial to low-achieving students? Evidence from Portugal. *Applied Economics, 50*, 4306–4317. https://doi.org/10.1080/00036846.2018.1444261.

Organisation for Economic Co-operation and Development (2003). *Student engagement at school: A sense of belonging and participation: Results from PISA 2000.* OECD Publishing. https://doi.org/10.1787/9789264018938-en.

Organisation for Economic Co-operation and Development (2020). *PISA 2018 results: Effective policies, successful schools* (Vol. 5). OECD Publishing. https://doi.org/10.1787/ca768d40-en.

Organisation for Economic Co-operation and Development (2019a). PISA 2018 assessment and analytical framework. OECD Publishing. https://doi.org/10.1787/b25efab8-en

Organisation for Economic Co-operation and Development (2019b). *PISA 2018 results: What school life means for students' lives* (Vol. 3). OECD Publishing. https://doi.org/10.1787/acd78851-en.

Organisation for Economic Co-operation and Development (2019c). *PISA 2018 results: Where all students can succeed* (Vol. 2). OECD Publishing. https://doi.org/10.1787/b5fd1b8f-en.

Peixoto, F., Monteiro, V., Mata, L., Sanches, C., Pipa, J., & Almeida, L. (2016). "To be or not to be retained…That's the question!" Retention, self-esteem, self-concept, achievement Goals, and grades". *Frontiers in Psychology, 7*, 1550. https://doi.org/10.3389/fpsyg.2016.01550.

Peixoto, F., Pipa, J., Mata, L., Monteiro, V., & Sanches, C. (2017). "To learn, or to be the best?" Achievement goal profiles in pre-adolescents. *Análise Psicológica, 35,* 499-511. https://doi.org/10.14417/ap.1182.

Pereira, M. C., & Reis, H. J. (2014). Grade retention during basic education in Portugal: Determinants and impact on student achievement. *Economic Bulletin, 16*, 61–83. https://www.bportugal.pt/sites/default/files/anexos/papers/ab201407_e.pdf

Pires, L., Santero-Sánchez, R., Macías, C. (2021). School failure in the region of Madrid (Spain): An approximation through diagnostic assessment in 2019. *Sustainability, 13*, 9895. https://doi.org/10.3390/su13179895.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696.

Santos, N., Monteiro, V., & Carvalho, C. (2022). Impact of grade retention and school engagement on student intentions to enroll in higher education in Portugal. *European Journal of Education*. https://doi.org/10.1111/ejed.12535.

Skaalvik, E. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology, 89*, 71-81. https://doi.org/10.1037/0022-0663.89.1.71

Slaten, C. D., Ferguson, J. K., Allen, K-A., Brodrick, D-V., & Waters, L. (2016). School belonging: A review of the history, current trends, and future directions. *Educational and Developmental Psychologist, 33*, 1-15. https://doi.org/10.1017/ edp.2016.6.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical Science, 25, 1–21 https://doi.org/10.1214/09-STS313.

Valbuena, J., Mediavilla, M., Choi, A., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys, 35*, 408-451. https://doi.org/10.1111/joes.12406.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1 - 67. https://doi.org/10.18637/jss.v045.i03.

Van Canegem, T., Van Houtte, M., & Demanet, J. (2021). Grade retention and academic self-concept: A multilevel analysis of the effects of schools' retention composition. *British Educational Research Journal, 47,* 1340-1360. https://doi.org/10.1002/berj.3729.

Van Canegem, T., Van Houtte, M., & Demanet, J. (2022). Grade retention: A pathway to solitude? A cross-national multilevel analysis of the effects of being retained on students' sense of belonging. *Comparative Education Review, 66*, 664-687. https://doi.org/10.1086/721649.

Wentzel, K. (2002). The contribution of social goal setting to children's school adjustment. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 222-246). Academic Press.

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology*, *102*, 135–152. https://doi.org/10.1037/a0016664.

Xiang, N., & Chiu, S. (2022). The school matters: Hong Kong secondary schools' grade-retention composition, students' educational performance, and educational inequality. *School Effectiveness and School Improvement*. https://doi.org/10.1080/09243453.2022.2136210.

5. STUDY 3: EFFECTS OF GRADE RETENTION IN LOWER SECONDARY EDUCATION ON STUDENTS' SELF-CONCEPT, SELF-ESTEEM, GOAL ORIENTATIONS, AND SCHOOL CAREER

This chapter is based on Pipa, J., Daniel, J. R., & Peixoto, F. (2023). *Effects of grade retention in lower secondary education on students' self-concept, self-esteem, goal orientations, and school career.* Manuscript submitted for publication. Results were also presented at 19th Biennial EARLI Conference (2021), and VII Encontro Nacional de Jovens Investigadores em Educação (2023).

Abstract

Grade retention is one of the most discussed and controversial educational measures, and yet, it is still widely applied in many countries. Grade retention has high educational, economic, and social costs, and thus, it is critical to expand the knowledge regarding its effectiveness. Research regarding the effects of grade retention on students' psychosocial development presents mixed findings. This is partly due to the nature of the variables considered, the grades assessed, and methodological issues, such as considering grade retention as a fixed measure and the lack of suitable comparison groups. This study aimed to analyse both the short to medium and longitudinal effects of grade retention in grades 7 or 8 on students': (1) academic self-concept, (2) self-esteem, (3) goal orientations and (4) school career. Data were collected continuously over a 3-year span (once a year) and, again, three years after the third wave. After matching 477 students on several pre-treatment variables using probability treatment weighting with time-varying treatments (i.e., retention), our analytical sample consisted of 85 promoted students, 33 students retained in grade 7, and 32 students retained in grade 8. Our results showed that retained students did not differ from their promoted peers in self-esteem and goal orientations in the short, medium, or long term. The exception was for an increase in 7th grade repeaters' academic self-concept in the years after retention. However, this advantage dissipated in the longer run. Finally, considering students' school career, grade retention was not predictive of further retention.

Keywords: academic self-concept, achievement goals, grade retention, matching methods

Introduction

In many countries, grade retention is the primary measure to ensure that low-achieving students acquire the academic knowledge and competencies required for a particular grade level (Allen et al., 2009). This is the case for European countries such as Belgium, Luxembourg, Germany, Spain, and Portugal, where more than 20% of students are retained at least once during their school career (Organisation for Economic Co-operation and Development [OECD], 2020). Despite its common use, recent reviews discourage the use of grade retention, considering its adverse effects on students' academic and non-academic outcomes (Allen et al., 2009; Goos et al., 2021; Valbuena, 2021). Nevertheless, the growing body of research addressing the effects of grade retention on psychosocial outcomes and school career still presents some controversial results due to the nature of the outcomes considered, the length of the studies, and the lack of high-quality methods employed to estimate its causal effects (Goos et al., 2021; Valbuena et al., 2021; Van Canegem et al., 2021).

Generally, methodological sound studies tend to compare the outcomes of retained students in a given grade with that of equivalent low-performing and promoted students. However, these studies have treated grade retention as a one-time fixed intervention instead of considering its dynamic time-varying nature, where students can be retained at different time points. As stated by some authors (Moser et al., 2012; Vandecandelaere et al., 2016), students who are promoted but are on the edge of being retained (i.e., at-risk students) are likely to be retained in the following grade(s). Thus, longitudinal studies that neglect the time-varying nature of grade retention may produce biased estimates of its effects since they compare retained students with (supposedly) continuously promoted students who were retained later (Vandecandelaere et al., 2016).

Therefore, in this study, we estimated grade retention effects using quasi-experimental methods that consider the time-varying nature of grade retention (Valbuena et al., 2021; Vandecandelaere et al., 2016). These effects were tested in 7th and 8th grade Portuguese students' academic self-concept, self-esteem, goal orientations, and school career during secondary education. These psychosocial and school career outcomes are understudied in the Portuguese context, highlighting the importance of this study.

Grade Retention Effectiveness

Grade retention, also referred to as grade repetition, requires that students are retained in grade, at the end of the school year, due to academic reasons (e.g., low achievement) and, thus, will have to repeat the same grade level in the next school year. The premise behind this practice is to give struggling students more time to master the learning material and to develop academic-related competencies (Hong & Yu, 2008; Smith & Shepard, 1988; Wu et al., 2010).

In addition, grade retention supposedly creates academically more homogeneous classrooms, which from teachers' viewpoint, eases instruction and educational support (Dupriez et al., 2008; Nunes et al., 2018; Smith & Shepard, 1988).

Opponents of grade retention, however, argue that the stigma associated with grade retention suppresses time benefits, especially in the longer run, by hindering students' competence beliefs, motivation, school engagement, and peer relationships (Goos et al., 2013; Mathys et al., 2019; Pagani et al. 2001). In addition, grade retention entails high economic and societal costs for the students themselves, the families, and the education system and society in broader terms (Ehmke et al., 2010; Goos et al., 2021; Valbuena et al., 2021).

Given the costs associated with grade retention, educators and researchers have sought to assess grade retention effectiveness by studying its academic and non-academic effects. More recent reviews of the research conducted during the past 20 years claim that the negative or 'no' effects of grade retention outweigh its potential benefits (Allen et al., 2009; Goos et al., 2021; Jimerson, 2001; Valbuena et al., 2021). In addition, these reviews showed that differential effects of grade retention could be found according to the: (a) context (e.g., country) where grade retention occurs, (b) outcome assessed (e.g., academic achievement vs. psychosocial and school career outcomes), (c) length of the study (short vs. long-run effects), and to the (d) methodology used (quasi-experimental vs naturalistic methods; Allen et al., 2009; Goos et al., 2021; Valbuena et al., 2021).

The Effects of Grade Retention on Psychosocial Outcomes

There are various reasons to expect grade retention effects on students' psychosocial outcomes, albeit in opposing directions. On the one hand, grade retention could positively impact students' self-beliefs and motivation (Marsh, 2016; Marsh et al., 2017). Retained students may finally experience success in academic-related tasks and receive positive teacher feedback, which could boost their motivation and learning behaviour (Ehmke et al., 2017; Goos et al., 2013).

Additionally, students may face grade retention as a new opportunity for a fresh start and would be more focused and motivated to master school-related material, as well as would be more competent and confident regarding academic subjects (Ehmke et al., 2017; Smith et al., 2022). This boost in motivation and self-beliefs is coupled with the advantaged conferred to retained students when placing them in a new classroom with younger students (Smith et al., 2022).

From a social comparison perspective, students use their group of classmates as a frame of reference to form their self-beliefs. In self-concept literature, this phenomenon is called the big-fish-little-pond effect (Marsh et al., 2008) and is often mentioned to explain the associations between students' grade retention and self-concept (Kretschmann et al., 2019;

Marsh, 2016; Marsh et al., 2017). Thus, by transiting to a classroom with younger and less experienced peers, retained students' self-perception of competence in specific academic domains (i.e., academic self-concept; Shavelson et al., 1976) would increase because they compare their academic accomplishments with that of their less academic experienced classmates (Marsh, 2016; Marsh et al., 2017; Marsh et al., 2008).

On the other hand, a substantial body of studies claims that grade retention could result in a negative psychological and emotional experience (Anderson et al., 2005; Jimerson & Ferguson, 2007; Martin, 2011), being identified by students as one of the major stressful life events (Anderson et al., 2005). Second, by repeating the exact same grade, students are confronted with the same learning material, either for the subjects they have failed or the ones they have succeeded. This mere rehearsal of all subjects may undermine students' motivation due to the deprivation of new and challenging tasks (Goos et al., 2013; Goos et al., 2021; Hong & Yu, 2008)

In addition, by being retained, students are detached from their peer group, which could trigger a feeling of having failed, and lowering students' motivation, engagement, and self-esteem (Goos et al., 2013; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Van Canegem et al., 2021; Xiang & Chiu, 2022). Transitioning to a class with younger students may also lead to lower acceptance of retained students by their new classmates (Demanet & Van Houtte, 2016; Pagani et al., 2001; Wu et al., 2010). These two interconnected experiences are consequences of the stigma associated with grade retention and the label of 'retainee' that students might receive from the school community (Wu et al., 2010; Xiang & Chiu, 2022).

Empirical studies considering students' psychosocial functioning after being retained have been less consistent, suggesting either the benefits or adverse effects of grade retention on these outcomes (Goos et al., 2021; Jimerson, 2001; Valbuena et al., 2021). These different results are due to the nature of the psychosocial variable considered (Goos et al., 2021; Martin, 2011; Van Canegem, 2021), the grade where retention occurred (Giano et al., 2022), whether the effects were assessed in short or longer term (Ehmke et al., 2010; Goos et al., 2013; Hwang & Cappella, 2018; Klapproth et al., 2016; Martin, 2011; Peixoto et al., 2016; Van Canegem et al., 2021; Wu et al., 2010), and whether a suitable comparison group of non-retained students was established (Hong & Yu, 2008; Martin, 2011).

In our study, we focused on grade retention effects in lower secondary education (i.e., grades 7 to 9), where the label of 'retainee' and the stigma associated with it may have more profound consequences (Demanet & Van Houtte, 2016; Giano et al., 2022; Mathys et al., 2019). Additionally, and because relatively less attention has been paid to the impact of grade retention on students' motivational orientations and self-esteem, despite playing a significant role in students' functioning and learning (Martin, 2009, 2011), we included these psychosocial outcomes in our study as well.

In Mathys et al. (2019) study, retained students' self-esteem, meaning an overall evaluation of satisfaction with oneself (Peixoto & Almeida, 2010), when compared to a matched group of promoted students, was negatively affected by grade retention immediately after being retained (i.e., at the beginning of the school year). Similarly, in a cross-sectional study, Martin (2011) found that students with a history of grade retention showed lower self-esteem levels in lower and upper secondary education compared to promoted students. Long-term adverse effects of grade retention on students' self-esteem were also found in the study of Hwang and Cappella (2018), who investigated the effects of grade retention in grades 1 and 2 at the age of 14. In the Portuguese context, the cross-sectional study of Nascimento and Peixoto (2012) of 9th-grade students, and the longitudinal study of Peixoto et al. (2016), considering the retrospective history of grade retention of two cohorts of students from 5th and 7th grades, found either no effects of grade retention on students' self-esteem (Nascimento & Peixoto, 2012) or adverse effects before and after the students were retained (Peixoto et al., 2016).

For academic self-concept in lower secondary education, the results from Nascimento and Peixoto (2012), and Peixoto et al. (2016) clearly showed that retained students scored lower than non-retained or low-achieving students. Similar results were found in cross-sectional studies investigating the impact of grade retention on students' academic self-concept in upper secondary education (Martin, 2011; Van Canegem et al., 2021). Longitudinal studies using a matched group of promoted students showed short-term benefits in grade 8 students' academic self-concept (Lamote et al., 2014) and math self-concept (Ehmke, 2010; Marsh et al., 2017), whereas in the study of Kretschmann et al. (2019) retained students in 6th grade revealed either lower German self-concept or equal levels of math self-concept when compared with promoted students during the retention year. However, these differential effects dissipated the year after retention (Klapproth et al., 2016; Kretschmann et al., 2019) and remained as students progressed (Lamote et al., 2014; Marsh et al., 2017).

The effects of grade retention on students' goal orientations, understood as the students' reasons for engaging in academic-related activities (Anderman & Johnson, 2002), have been so far understudied. However, the few studies addressing this motivational approach systematically showed that retained students presented lower levels of task orientation after being retained (Martin, 2009; Nascimento & Peixoto, 2012; Peixoto et al., 2016; Peixoto et al., 2017). In addition, Kretschmann et al.'s (2019) study showed that students are less task-oriented even before being retained. Task orientation is considered the most valuable goal orientation since it is defined by the goal of engaging in an academic task for learning and gaining competence instead of an external reward (Anderman & Johnson, 2002; Skaalvik, 1997), and it is closely linked with achievement (Kaplan & Maher, 2007; Skaalvik, 1997). In line with this, retained students presented higher levels of avoidance orientation

(Nascimento & Peixoto, 2012; Peixoto et al., 2016), meaning that retained students are more oriented to avoiding learning situations (Skaalvik, 1997). Regarding the goal orientations of engaging in learning activities to outperform others and to demonstrate competence over others (*i.e.*, ego orientations; Anderman & Johnson, 2002; Kaplan & Maher, 2007; Skaalvik, 1997), studies have shown no differences between retained and promoted students either before (Kretschmann et al., 2019) or after being retained (Nascimento & Peixoto, 2012; Peixoto et al., 2016).

The Effects of Grade Retention on Students' School Career

Investigating the impact of grade retention on students' later school career provides critical information regarding the long-lasting effects of grade retention. It contributes to unravelling its economic and social costs. Retained students have often revealed negative educational trajectories (Goos et al., 2021; Valbuena et al., 2021). Particularly, grade retention has been consistently associated with high-school dropout, being among one of its highest risk factors (Gubbels et al., 2019; Mendez et al., 2015).

The effects of grade retention on later retention have been investigated to a lesser extent. Generally, it is believed that teachers tend to retain students only once to avoid exacerbating the age gap between retained students and their classmates (Moser et al., 2012; Nunes et al., 2018). Nevertheless, although it may apply to primary education (Moser et al., 2012; Goos et al., 2013; Nunes et al., 2018), some studies showed that grade retention is associated with an increased likelihood of being retained in secondary education (Geng & Rockoff, 2017; Mendes et al., 2015).

Grade Retention in Portugal

The Portuguese education system comprises 12 years of compulsory education (Eurydice, 2022) and is organised in basic (grades 1 to 9) and secondary education (grades 10 to 12). Basic education is divided into three consecutive cycles: the first cycle lasts from grades 1 to 4, the second from grades 5 and 6, and the third from grades 7 to 9. Basic education is uniform in structure (Dupriez et al., 2008) since these cycles share a common core curriculum, whereas secondary education requires vocational choices (European Commission, 2020).

According to Dupriez et al. (2008), educational systems with a uniform structure face greater challenges in dealing with students' heterogeneity in terms of academic competencies. Thus, one of the strategies used more often is grade retention. The Portuguese legislation presently states that grade retention should be only used as an 'exceptional measure' when promoting the student to the next grade compromises the acquisition of new learnings (Decree-Law no. 55/2018, 2018). The current legislation is similar to that in force at the start of the data

collection for this study (2012/2013 school year) (Conselho Nacional de Educação [CNE], 2015). At that time, grade retention was already considered an exceptional measure with a pedagogical purpose, and its application was mainly based on students' achievement: students were retained when their grade marks were below 3 (on a scale from 1 to 5) in three school subjects (e.g., Portuguese, Math, and another compulsory subject; CNE, 2015). Despite this recommendation, grade retention was a relatively common practice, especially in the third cycle of basic education, where the grade retention rates were around 16% during the school year 2012/2013 (CNE, 2020). However, most recent statistics show a downward trend, with a 6% rate recorded for the 2018/2019 school year. Considering the international context, recent data shows that Portuguese retention rates are still among the highest of the European countries (OECD, 2020): in the 2018 cycle of the Programme for International Student Assessment (PISA), 27% of 15-year-old students reported having been retained at least once during their school career. A number largely above the OECD average of 11% (European Commission, 2020; OECD, 2020).

In summary, despite the exceptionality feature ascribed to grade retention and the downward tendency, schools and teachers continue to rely on grade retention as a tool to deal with students' heterogeneity and low achievement. In addition, these rates may reflect a positive belief shared by educators of the benefits of grade retention on the one hand and the difficulty in finding alternative tools and measures to avoid grade retention on the other (European Commission, 2020; Goos et al., 2021; Santos et al., 2023).

The Present Study

Considering the literature review presented above, this study was grounded on three primary considerations: a) the established relationship between grade retention and psychosocial variables; b) the higher retention rates observed in the Portuguese context; and c) the need for studies using high-quality methods to model grade retention effects, especially using Portuguese data. Therefore, we investigated grade retention effects, considering its time-varying nature (Vandecandelaere et al., 2016), on the development of academic self-concept, self-esteem, and goal orientations during lower secondary education (i.e., grades 7 to 9), for students who were: (a) continuously promoted, (b) retained in 7th grade, or (c) retained in 8th grade.

Secondly, and using a smaller sample of students who were followed during their secondary education (i.e., from 9th grade to 12th grade), we aimed to investigate the longitudinal effects of grade retention, considering students' school career through subsequent grade retention. In addition, we explored the long-term effects of grade retention, assessing the psychosocial outcomes three years after the third wave of data collection. This aim was considered only in exploratory terms, given the small number of observations.

We expected that, on average, Portuguese retained students would not benefit from grade retention, based on previous literature using a similar methodological approach in different samples (Hwang & Cappella, 2018; Klapproth et al., 2016; Kretschmann et al., 2019; Lamote et al., 2014). More specifically, we expected that students would decline their self-concept, self-esteem and motivation before being retained. This decline would persist or stabilise over the years, over and above the eventual short-term gains explained by the big-fish-little pond effect. In addition, we expected detrimental effects of grade retention on retained students' school career through subsequent retention, as previously described (Geng & Rockoff, 2017; Mendes et al., 2015).

Method

Participants

The data of the current study were taken from a 3-year longitudinal research project (2012/2013–2014/2015), studying the development of affective components of learning in two cohorts of 1807 middle school students, attending the 5th and 7th grade (at the beginning of the project), in 12 schools in the Lisbon region. Data were collected once every year in the middle of the school year (March-May), resulting in three waves of data collection (Y1 to Y3). Four hundred and forty-eight students were excluded from this study either because they were absent on the first data collection or had already repeated a year in school, which would have challenged the disentanglement of grade retention effects.

For this study, we relied on data from the 7th grade students' cohort (N = 552). Students who were absent during the second and third measurement points or moved to a different school were considered dropouts and thus excluded from the analysis (n = 75). As a result, the final sample comprised 477 students, aged 12.35 on average (*SD* = 0.55) in Y1 (7th grade), of whom 48% were boys. Regarding grade retention, 33 students were retained in 7th grade, and 40 were retained in 8th grade. No differences between dropout and participating students were found in most background variables (i.e., gender, parental education, and school context) except for students' age, $t(545) = 2.71$, $p = 0.007$. These differences are, however, small: the mean age for dropout students was 13.56 and for participating students was 13.35, as mentioned above.

Additionally, in the 2017/2018 school year, a smaller study with a new wave of data collection (W4) was conducted when students were between the 10th and 12th grades (i.e., three years after Y3). Data in all four measurement points were available for 31 students (11 retained; 18 boys) attending five schools. This low participation rate may be attributed to the time gap between Y3 and W4 data collection points and student mobility since students often change schools after the 9th grade. As such, measurements for this data point were only analysed in exploratory terms.

Measures

Independent Variable

Grade Retention. Grade retention was obtained every year from school records and was operationalised as 0 = student was not retained and 1 = student was retained. Based on this information, for the outcome analyses, students were assigned to one of three groups (1 = never retained, 2 = retained in 7th grade, 3 = retained in 8th grade).

Outcome Variables

Self-concept and Self-esteem. Students' academic self-concept, self-esteem and importance attributed to academic self-concept were measured using the Self-concept and Self-esteem Scale for Adolescents (Peixoto & Almeida, 2010). This scale consists of 51 items distributed across ten subscales, assessing nine specific academic and non-academic domains of self-concept and self-esteem. Academic self-concept was derived from three dimensions, where students were asked to rate, on a 4-point response scale, their: (a) academic competence (e.g., "Some students are fast in doing their schoolwork), (b) language competence (e.g., "Some students find it very easy to write"), and (c) math competence (e.g., "Some students feel they are good at math"). Self-esteem (e.g., "Some students are happy with themselves most of the time") and importance attributed to academic self-concept (e.g., "Some students believe that it is important to be a good student") were also assessed using a 4-point response scale. Scores in each subscale were averaged; thus, higher scores refer to higher academic self-concept, self-esteem and importance attributed to academic self-concept. Reliability, assessed through Cronbach's alpha, showed adequate internal consistency in all dimensions and all data collection points, ranging from α = 0.78 for importance attributed to academic self-concept in Y1 and α = 0.87 for academic self-concept in Y3.

Goal Orientations. Students' achievement goal orientations were assessed using the Goal Orientations Scale (Pipa et al., 2016), measuring through 24 items and on a 4-point response scale, four types of goal orientations within the academic context: (a) task orientation (e.g., "For some students, it is important to learn new things at school"), (b) self-enhancing ego orientation (e.g., "Some students always try to do better than their classmates"), (c) self-defeating ego orientation (e.g., "In class, some students try not to make a fool of themselves when the teacher asks questions"), and (d) avoidance orientation (e.g., "At school, some students like to do as little as possible"). Cronbach's alpha ranged from 0.78 in task orientation at Y1 and 0.91 in self-defeating orientation at Y3, revealing good internal consistency of these measures. Scores on each dimension were also averaged, and higher scores represented higher levels in each goal orientation.

Covariates

Students' Gender. Gender was operationalised as a dichotomous variable where 0 = female and 1= male.

Students' Age. Students provided their age at the time of data collection.

Parental Education. Students were asked to provide information regarding their mother and father's highest level of education. Parental education resulted from averaging both scores.

Students' Achievement. We used information regarding students' school marks in Portuguese and Math subjects each year. School marks in each school term were obtained from students' records and averaged, ranging between 1 and 5.

School Retention Composition. The proportion of existing retained students in each school was obtained through official records every year (Ministério da Educação/Direcção Geral de Educação, n.d.).

Students' Math Reasoning. In Y1, students' math reasoning was measured by asking the students to complete a series of 24 sequences (Almeida & Lemos, 2007). Students received 1 point for each correct answer, and scores were summed to obtain an index for math reasoning. The internal consistency for this measure was $\alpha = 0.89$.

Procedures

Data Collection

Schools from the Lisbon area were contacted to be part of the research project. The project and its procedures were reviewed and approved by the Directorate-General of Education. Parental consent was required for participation, and students' assent was obtained. Students were informed about the aims of the study, and confidentiality was assured. Trained researchers administered the outcome measures in a classroom session and other measures relevant to the broader study. Data collection occurred every year, between March and May, except for grade retention and achievement measures, which were retrieved from students' records during and at the final of the school year.

For collecting the fourth measurement point, schools were again contacted and asked to participate in this data collection process. A new parental consent was sent to students' parents, and assent from participating students was obtained. This smaller study was also approved by the Directorate-General of Education and, internally, by the Institutional Ethical Committee (D/011/01/2019).

Data Analysis

Handling Missing Data. The mean level of missing information in the original sample for both covariates and outcome variables was 2.5%, ranging from 0.42% to 8%. To overcome

this issue, and since the propensity score technique requires completed datasets, we imputed missing information through Multiple Imputation by Chained Equation Modelling using the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011). Using this procedure, we generated five complete datasets (10 predictive model iterations per dataset). Subsequent analyses were executed in each imputed dataset, and estimates were then pooled (Rubin, 1987).

Group Comparison Strategy. In this study, we aimed to estimate the effects of an intervention: grade retention. This is particularly challenging for ethical reasons since randomly assigning students to grade retention, and a promotion group is unfeasible. Thus, to counter selection bias and produce reliable estimates for treatment effects, researchers have recommended using quasi-experimental methods over regression modelling (Allen et al., 2009; Austin, 2011; Goos et al., 2021). This approach is critical when the treatment (i.e., retention) is most likely applied to students with low achievement characteristics (Allen et al., 2009; Austin, 2011; Goos et al., 2021; Kretschmann et al., 2019). Following this recommendation, we used propensity score methods (Austin, 2011) to reduce bias between retained and promoted students in pre-treatment covariates.

Among the different propensity score methods, we chose inverse probability treatment weighting with time-varying treatments (Austin & Stuart, 2015; Vandecandelaere et al., 2016), given that (as the name of the method implies) it allows considering the time-varying nature of the treatment- that is, the fact that students could be retained at two-time points. This method also potentiates the use of all observations by weighting subjects according to the group that they represent, where treatment subjects receive a weight of 1 and control subjects are weighted by the inverse of the probability of receiving the treatment (Austin & Stuart, 2015; van der Wal & Geskus, 2011). Balance in weighting treated and comparison subjects across covariates was achieved when: (a) standardised mean differences in the covariates were below 0.25, (b) when variance ratios were between 0 and 2, and (c) graphical inspection revealed no major concerns (Stuart, 2010; Vandecandelaere et al., 2016). Since students were retained at different time points, the balance between the groups was checked at Y1 for promoted and 7th grade repeaters, and at Y1 and Y2 for promoted and 8th grade repeaters (Vandecandelaere et al., 2016).

Regarding the covariates used to assess the balance between treated and comparison subjects, we choose variables that were either associated with the treatment or the outcomes (Allen et al., 2009; Kretschmann et al., 2019). Specifically, we adjusted for differences in 14 variables, including students' background variables (i.e., age, gender, parental highest educational level, math reasoning), school characteristics (i.e., school retention composition), and baseline measures in the outcome variables. The selection of these variables was

theoretically and empirically based (e.g., Demanet & Van Houtte, 2016; Klapproth & Schaltz, 2015; Kretschmann et al., 2019; Nunes et al., 2018; Pereira & Reis, 2014).

The steps for establishing the comparison group were based on Vandecandelaere et al. (2016) and were as follows. First, we fit a logistic regression estimating the probability of being retained at Y1 as a function of pre-treatment covariates. This step allowed us to consider only at-risk students with some probability of being retained and avoid overweighting (Austin & Stuart, 2015). Thus, we restricted our analytic sample to those students with at least a 1% probability of being retained. We also inspected the results by increasing the cut-off probability. However, they remained virtually the same whether we used a 1% or 5% probability of being retained, as per Vandecandelaere et al. (2016). After, we applied the inverse probability treatment weighting using the ipw package in R (van der Wal & Geskus, 2011) and assessed the balance between the groups, as previously described, using the probabilities obtained from the logistic regression. The weights obtained were considered in the outcome analyses.

Outcome Analysis. Outcome analyses were based on a same-age comparison, where retained students were compared to their 7th grade classmates who were continuously promoted. To estimate the effects of grade retention on the outcomes, we intended to use multilevel models, given the multilevel nature of the data, where students are nested within classes and schools (Marsh, 2016). However, due to the number of observations in the different levels of the data (students and schools), it was not warranted to use multilevel models. Inspection of the intraclass correlation coefficient revealed low variance between schools, ranging between 0 and 10% for the different outcomes. As such, a weighted regression model was computed for each outcome, with the wave (i.e., Y1 to Y3) and the students grouping variable (i.e., promoted students, 7th grade and 8th grade repeaters) as categorical variables, as well as the interaction between both. In addition, we followed Stuart's (2010) recommendation and incorporated variables in the regression models used in the weighting procedure to reduce the minor differences that could have remained after the weighting. All variables were grand mean-centred prior to modelling.

Considering W4 outcomes, we conducted a logistic regression model to explore students' school career (i.e., subsequent grade retention) and one regression model for each outcome variable to estimate the long-term effects of grade retention. Given the limited number of observations, we did not include any additional covariates in these models.

Results

Group Balance

Before conducting the outcome analysis, selection bias in the background variables and baseline measures of the outcomes was inspected through inverse probability treatment weighting with time-varying treatments. The analyses showed that a better balance would be

achieved after applying the cut-off of 1% probability of being retained, reducing our observations to 85 promoted students, 33 retained students in 7th grade and 32 retained students in 8th grade. Our analytical sample was then comprised of 150 retained and promoted students. Table 1 presents descriptive statistics for the background variables at Y1 and the outcome variables during the first three years of data collection (Y1 to Y3). Table 2 shows the standardised mean differences (SMD) between the promoted and retention groups.

Considering pre-retention (i.e., Y1) differences between promoted students and retained students in 7th grade, larger differences before weighting in the background variables were observed in parental education (*d* = -0.87), math reasoning (*d* = -0.80), Portuguese achievement (*d* = -1.75), and math achievement (*d* = -1.56). After weighting, the SMD was successfully reduced in most variables, ranging between 0.05 and 0.21. The exception were parental education (d = 0.39) and students' age (*d* = -0.30), which remained imbalanced after weighting. Regarding baseline measures in the outcome variables before weighting, the major differences were found in academic self-concept (*d* = -1.08) and task orientation (*d* = -0.52). After weighting, these differences and the differences in the remaining baseline variables were substantially reduced, with all SMD below 0.25.

Comparing promoted and retained students in 8th grade achieved virtually the same results. In parental education, the SMD before weighting (*d* = -0.64 in Y1 and *d* = -0.65 in Y2) remained imbalanced after weighting in Y1 (*d* = 0.61), but in Y2, the balance was achieved (d = -0.05). For Portuguese achievement, the differences also remained slightly imbalanced after weighting (*d* = 0.33 for Y1 and *d* = 0.39 for Y2). However, it should be mentioned that before weighting, the differences were above 1; thus, we reduced these differences substantially with weighting. For math reasoning and math achievement, the SMD before weighting was considerably large (*d* = -0.52 in Y1 and Y2 for math reasoning; *d* = -1.42 in Y1 and *d* = -1.28 in Y2 for math achievement). Weighting allowed us to reduce these differences successfully (*d* = 0.25 in Y1 and *d* =-0.12 in Y2 for math reasoning; *d* = 0.04 in Y1 and *d* = 0.05 in Y2 for math achievement). In the remaining background variables, the differences before weighing were already below 0.25 and remained virtually the same or smaller after weighing. Finally, considering the outcome measures, larger SMD before weighting were found in academic self-concept (*d* = -0.81 in Y1 and *d* =-1.19 in Y2), in task orientation (*d* = -0.59 in Y2) and self-enhancing orientation (*d* = 0.34 in Y1). After weighting, all these SMD were below 0.25 (*d* = 0.23 in academic self-concept; *d* = 0.10 in task orientation; *d* = -0.08 in self-enhancing orientation). Differences in the other outcome measures were already below 0.25 before weighting and remained below that threshold.

Table 1

Descriptive Statistics for the Variables Under Study

Variable		Promoted		Retained 7th		Retained 8th	
	Year	*M*	*SD*	*M*	*SD*	*M*	*SD*
Background variables Y1							
Boys		48%		52%		56%	
Age		12.40	0.63	12.60	0.75	2.30	0.47
Parental education		10.30	2.78	8.45	2.76	8.03	3.50
Math reasoning		13.30	4.83	11.70	4.79	11.80	4.58
School retention composition		26.70	11.1	27.80	10.6	29.80	12.3
Portuguese achievement		2.49	0.53	2.21	0.42	2.34	0.48
Math achievement		2.24	0.48	2.15	0.51	2.22	0.42
Outcome variables							
Academic self-concept	1	2.11	0.35	2.00	0.42	2.05	0.39
	2	2.12	0.38	2.19	0.34	1.93	0.29
	3	2.13	0.36	2.27	0.36	2.10	0.31
Self-esteem	1	2.48	0.55	2.40	0.48	2.40	0.46
	2	2.56	0.44	2.55	0.42	2.50	0.40
	3	2.52	0.46	2.56	0.38	2.53	0.43
Importance academic	1	2.20	0.26	2.15	0.28	2.24	0.21
self-concept	2	2.12	0.26	2.12	0.23	2.16	0.20
	3	2.08	0.26	2.06	0.21	2.10	0.25
Task orientation	1	2.67	0.37	2.56	0.37	2.77	0.36
	2	2.60	0.38	2.55	0.38	2.52	0.29
	3	2.53	0.36	2.51	0.34	2.54	0.37
Self-enhancing orientation	1	2.39	0.53	2.30	0.53	2.40	0.48
	2	2.17	0.49	2.12	0.60	2.20	0.45
	3	2.19	0.47	2.18	0.48	2.25	0.61
Self-defeating orientation	1	2.13	0.60	2.16	0.62	2.19	0.62
	2	2.05	0.61	1.88	0.56	2.02	0.64
	3	1.99	0.65	1.98	0.63	1.98	0.60
Avoidance orientation	1	2.14	0.50	2.13	0.51	2.04	0.46
	2	2.25	0.50	2.08	0.48	2.25	0.38
	3	2.32	0.48	2.27	0.53	2.28	0.38

Table 2

Standardized Mean Differences Before and After Weighting

	Promoted[a] Vs.Retained 7th[b] Y1		Promoted Vs. Retained 8th[c] Y1		Promoted Vs. Retained 8th[c] Y2	
	Before weighting	After weighting	Before weighting	After weighting	Before weighting	After weighting
Boys	0.08	0.08	0.10	-0.16	0.10	-0.00
Age	0.35	-0.30	-0.15	0.13	-0.17	0.14
Parental education	-0.87	0.39	-0.64	0.61	-0.65	-0.05
Math reasoning	-0.80	0.09	-0.52	0.25	-0.52	-0.12
School retentioncomposition	0.01	-0.17	0.23	-0.29	0.23	-0.26
Portuguese achievment	-1.75	0.21	-1.07	0.33	-1.40	0.39
Math achievement	-1.56	0.05	-1.42	0.04	-1.28	0.05
Academic self-concept	-1.08	0.08	-0.81	0.23	-1.19	0.23
Self-esteem	-0.33	-0.02	-0.24	0.20	-0.14	0.12
Importance academic self-concep	-0.48	0.25	-0.01	-0.18	-0.16	-0.10
Task orientation	-0.52	0.10	0.01	-0.16	-0.59	0.10
Self-enhancing orientation	-0.18	0.16	0.06	-0.10	-0.08	-0.27
Self-defeating orientation	0.13	0.08	0.34	-0.08	0.17	-0.05
Avoidance orientation	0.26	0.09	-0.04	0.14	0.40	-0.09

Note. [a]*n* = 85, [b]*n* = 33, [c]*n* = 32. Negative standardized mean differences indicate lower scores for retained students

Grade Retention Effects

Concerning our first aim of investigating the effects of grade retention on the assessed psychosocial variables over three school years, several regression models were computed. The results are presented in Table 3 and illustrated in Figure 1. First, considering students' self-concept and self-esteem development, the results showed that promoted students presented higher academic self-concept levels before any retention groups were retained (Y1). Nonetheless, we observed an increase in academic self-concept for students retained in 7th grade over the years, reaching statical significance in Y3 (b = 0.20, SE = 0.09, p = 0.030). As Figure 1a shows, this group of students substantially increased their perception of academic competence during the retention year (Y2) and maintained this trajectory the following year. The group of retained students in 8th grade presented lower levels of academic self-concept either before or after being retained (see Figure 1a); however, these differences were not statistically significant. For self-esteem, promoted and retained students presented similar and stable trajectories across the years (see Figure 1b). Considering students' importance attributed to academic self-concept, a slight decline was observed for promoted students over the years (b = -0.09, SE = 0.04, p = 0.004 for Y2, and b = -0.10, SE = 0.04, p = 0.004 for Y3). For the retention groups, the results were not statistically significant, yet, as Figure 1c shows, these groups of students also presented downward trajectories, which seemed to stabilise after retention occurred (i.e., from Y2 onwards).

Concerning students' development of goal orientations, all the groups have declined their task orientation from Y1 to Y2 (see Figure 1d), but this was only significant for promoted students (b = -0.09, SE = 0.04, p = 0.043). After Y2, the task orientation of promoted and retained students in 8th grade remained stable. In contrast, retained students in 7th grade slightly declined their task orientation in the year after the retention year (Y3), as shown in Figure 1d, although these differences were not significant. For self-enhancing orientation, a steeper decline was found from Y1 to Y2, but again, this decline was only statistically significant for promoted students (b = -0.26, SE = 0.06, p < 0.001). In Y3, both retained students in 7th grade and promoted students slightly recovered their levels of self-enhancing orientation, but without reaching Y1 levels. This effect was, however, only significant for promoted students (b = -0.16, SE = 0.06, p = 0.015; see Figure 1e). Considering self-defeating orientation, promoted students were engaged in a downward trajectory from Y1, reaching significance levels at Y3 (b = -0.17, SE = 0.03, p = 0.004). Retained students in 8th grade presented higher levels of this orientation; however, they also engaged in a downward trajectory during the year of the retention decision and the year after. A steeper decline was observed in retained students in 7th grade (see Figure 1f) from the year of the retention decision (Y1) to the retention year (Y2). Moreover, this group of students slightly increase their self-defeating orientation in Y3, reaching the same levels as the other groups. These differences were, nonetheless,

nonsignificant. Finally, for avoidance orientation, a similar upward trajectory was found for promoted and retained students in 8th grade from Y1 to Y3, reaching significance in Y3 for promoted students (b = 0.13, SE = 0.05, p = 0.013). In addition, avoidance orientation of retained students in 7th grade was maintained during the retention year; however, in the year after, a steeper increase, although nonsignificant, was observed for this group, as shown in Figure 1g.

Figure 1

Graphical Illustration of the Development of the Outcomes over Time for Promoted Students, Retained Students in 7th Grade and Retained Students in 8th Grade

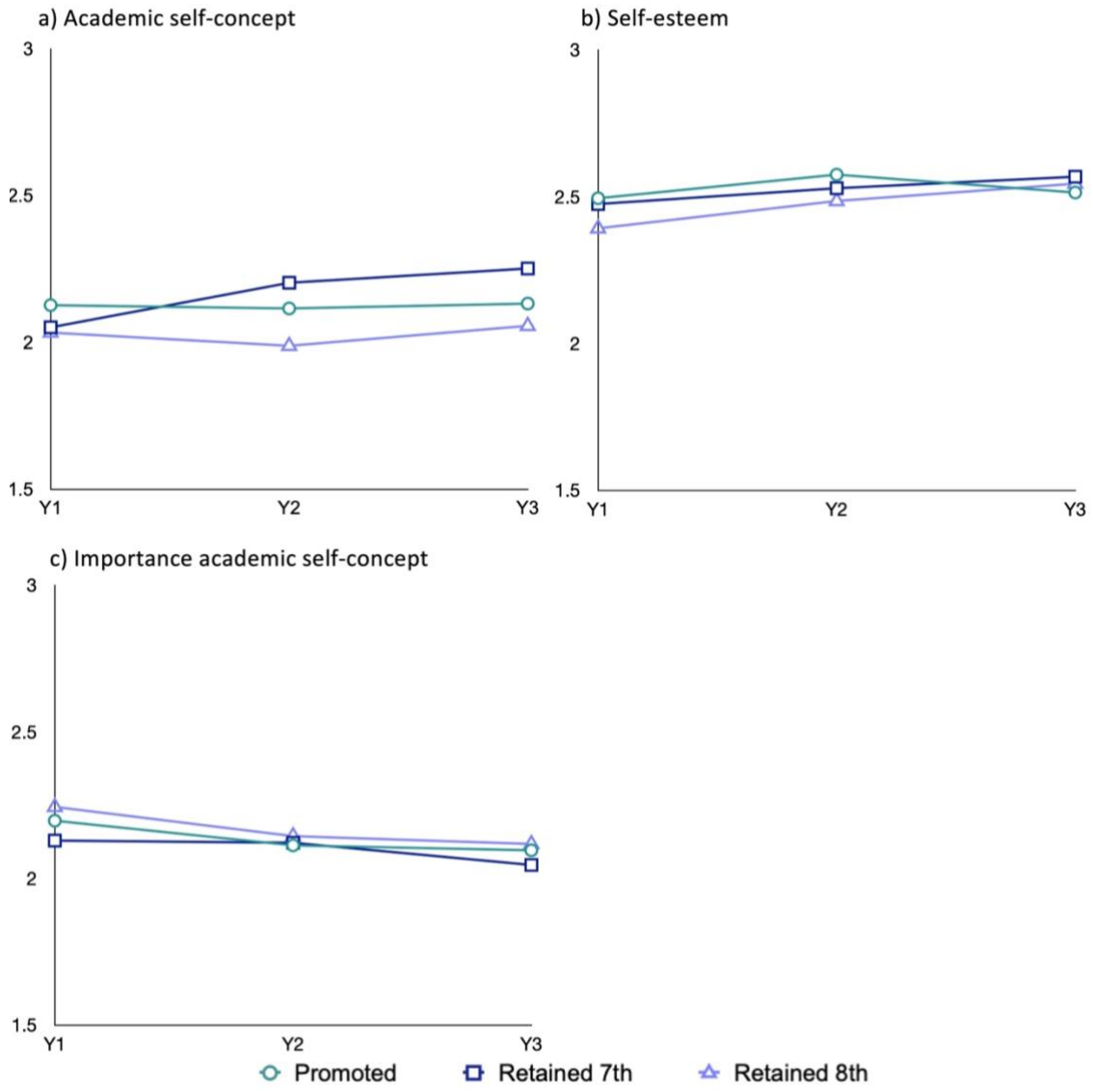


Figure 1 (continued.)

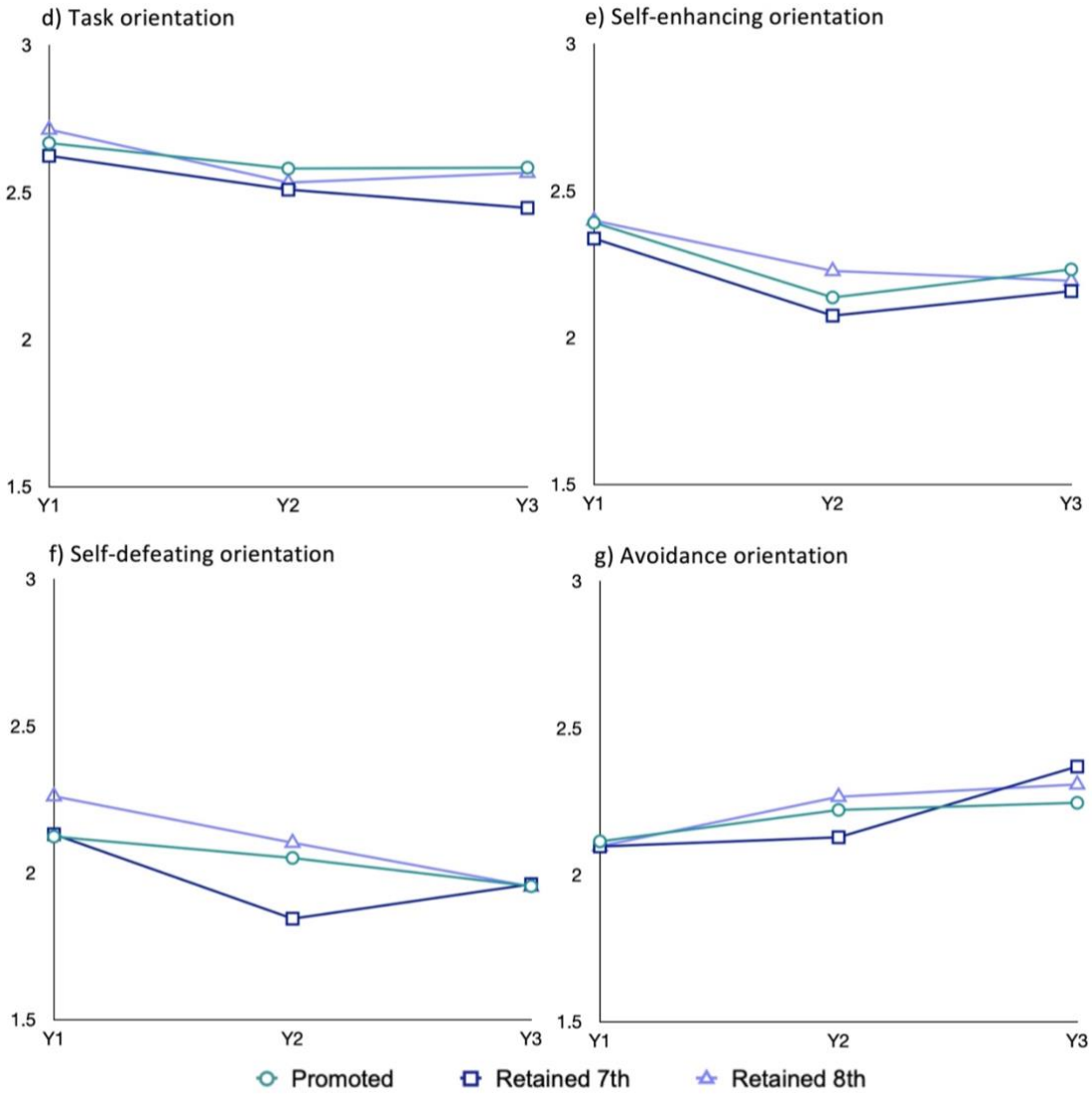


Table 3

Regression Coefficients of Grade Retention on Psychosocial Outcomes

	Promoted			Retained 7th			Retained 8th		
	Y1	Y2	Y3	Y1	Y2	Y3	Y1	Y2	Y3
Academic self-concept	2.13***	-0.01	0.01	-0.08	0.16	0.20*	-0.09	-0.03	-0.02
	(0.06)	(0.05)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)	(0.08)	(0.08)
Self-esteem	2.50***	0.08	0.02	-0.02	-0.03	0.07	-0.10	0.01	0.13
	(0.05)	(0.05)	(0.06)	(0.10)	(0.11)	(0.13)	(0.10)	(0.10)	(0.10)
Importance self-concept	2.20***	-0.08**	-0.10**	-0.07	0.08	0.02	0.05	-0.02	-0.03
	(0.03)	(0.03)	(0.03)	(0.06)	(0.06)	(0.07)	(0.05)	(0.06)	(0.06)
Task orientation	2.67***	-0.09*	-0.08	-0.04	-0.03	-0.09	0.05	-0.09	-0.06
	(0.04)	(0.04)	(0.06)	(0.09)	(0.10)	(0.11)	(0.08)	(0.09)	(0.10)
Self-enhancing orientation	2.40***	-0.25***	-0.16*	-0.05	-0.01	-0.02	0.01	0.08	-0.06
	(0.05)	(0.06)	(0.06)	(0.12)	(0.13)	(0.15)	(0.10)	(0.11)	(0.12)
Self-defeating orientation	2.12***	-0.07	-0.17**	0.01	-0.22	-0.01	0.14	-0.09	-0.14
	(0.05)	(0.05)	(0.06)	(0.14)	(0.16)	(0.16)	(0.11)	(0.12)	(0.12)
Avoidance orientation	2.12***	0.11	0.13	-0.02	-0.08	0.14	-0.02	0.06	0.08
	(0.05)	(0.05)	(0.05)	(0.12)	(0.11)	(0.13)	(0.10)	(0.11)	(0.10)

Note. Standard errors are in parenthesis. Y1 = 7th-grade, the year of retention decision for 7th-grade repeaters; Y2 = 8th-grade for promoted students, the retention year for 7th-grade repeaters, the year of retention decision for 8th-grade repeaters; Y3 = 9th-grade for promoted students, the year after the retention year for 7th-grade repeaters, the retention year for 8th-grade repeaters

*p < .05, ***p < .001, **p < .01

Concerning our second exploratory goal of investigating the longitudinal effects of grade retention, the results regarding students' school career showed that being retained did not increase the likelihood of subsequent grade retention (*b* = 1.15, *SE* = 0.94, *p* = 0.220). In addition, considering the long-term effects of grade retention, as seen in Table 4, regression analyses showed nonsignificant effects of grade retention on W4 outcomes despite the negative estimates observed.

Table 4

Means, Standard Deviations, and Regression Coefficients for the Longitudinal (W4) Effects of Grade Retention

	Promoted Students [a]		*Retained students* [b]		
	M	*SD*	*M*	*SD*	*b*
Academic self-concept	2.96	0.49	2.68	0.35	-0.28 (0.17)
Self-esteem	3.08	0.54	2.77	0.65	-0.30 (0.22)
Importance academic self-concept	3.30	0.72	3.12	0.35	-0.18 (0.23)
Task orientation	3.38	0.48	3.18	0.48	-0.19 (0.18)
Self-enhancing orientation	2.39	0.75	2.25	0.35	-0.14 (0.24)
Self-defeating orientation	1.78	0.67	2.27	0.98	0.49 (0.23)
Avoidance orientation	2.02	0.66	2.21	0.51	0.20 (0.23)

Note. [a] n = 20, [b] n=11, standard errors are in parenthesis

Discussion

Effects of Grade Retention on Students' Psychosocial and School Career Outcomes

This study investigated the effects of grade retention in lower secondary education on psychosocial outcomes across three school years by employing matching procedures through inverse probability treatment weighting with time-varying treatments (Austin & Stuart, 2015; Vandecandelaere et al., 2016). Our approach allowed us to reduce pre-retention differences between promoted and retained students and followed the recommendations of previous meta-analysis (Allen et al., 2009; Goos et al., 2021) of using quasi-experimental methods to provide more solid conclusions regarding grade retention effectiveness. Moreover, only a few studies considered the natural history of grade retention (Vandecandelaere et al., 2016) by estimating grade retention effects using time-varying estimates and, as such, this study adds to the body of literature claiming the importance of considering the time varying-nature of grade retention (Moser et al., 2012; Valbuena et al., 2021; Vandecandelaere et al., 2026). In addition,

our waves of data collection allowed us to investigate the effects on students' self-concept, self-esteem and goal orientations after retention has occurred and to consider the period before retention, as suggested by Kretschmann et al. (2019). Finally, this study also explored the long-term effects of grade retention on students' psychosocial variables and school career, considering students' further grade retention. The design of this study, along with its outcomes, remained understudied in Portugal, a country with a considerable culture of grade retention, especially in secondary education (CNE, 2020).

Concerning the results before retention, i.e., Y1 for 7th grade repeaters and Y2 for 8th grade repeaters, our results challenge the idea that the threat of grade retention would motivate students to work harder (Belot & Vandenberghe, 2014). As predicted, retained students scored lower on most of the self-perception and motivational variables before being retained, corroborating the findings from Kretschmann et al. (2019) study, which found more significant differences between retained and promoted students before retention occurred. In addition, as posited by Kretschmann et al. (2019), these findings reveal that students, by anticipating the possibility of grade retention, may engage in negative self-judgments, including the feeling of failure, even before being retained. This finding adds to previous studies, which stated that students could experience feelings of having failed after being retained (Alexander et al., 2003; Goos et al., 2013; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Van Canegem et al., 2021; Xiang & Chiu, 2022).

As hypothesized, the results referring to the trajectories of students' self-esteem and goal orientations did not reveal statistical differences between retained students and their at-risk promoted peers. Instead, in most of these outcomes, both retained and promoted students showed similar trajectories across three school years. These findings support the results from previous studies using matching methods (e.g., Hwang & Cappella, 2018; Klapproth et al., 2016; Kretschmann et al., 2019) providing little support for the 'gift of time' idea in respect of grade retention. Together, these results suggest that retained students do not benefit from grade retention in the sense that they would have developed equally their self-esteem and motivational orientations had they been promoted instead of being retained (Hwang & Cappella, 2018; Klapproth et al., 2016; Kretschmann et al., 2019; Lamote et al., 2014; Martin, 2011). Additionally, we observed either a downward or a low and stable trajectory in self-esteem, the importance attributed to academic self-concept, and goal orientations for both groups of students. These results, along with the upward trajectory in avoidance orientation, suggest that students may continuously devalue academic-related tasks to avoid failure and protect their feelings of self-worth (Nascimento & Peixoto, 2012; Peixoto et al., 2016; Peixoto et al., 2017).

Conversely, our study showed a positive grade retention effect on students' self-concept for the group of 7th grade repeaters. This result could be explained by the big-fish-

little-pond effect, in which retained students' self-perception of competence benefits from grade retention (Marsh, 2016; Marsh et al., 2017; Marsh et al., 2008) because of the change of reference group to younger and less experienced classmates. On the other hand, the group of 8th grade repeaters showed consistently lower levels of academic self-concept. One explanation for these differences may be that 7th grade repeaters attribute their experience of unsuccess (grade retention) to external causes, i.e., the transition to a new cycle, which might attenuate the adverse effects of grade retention. However, it is worth noting that for all the variables considered, the scores of both groups of students are considerably low, especially in comparison with previous studies using similar Portuguese samples (e.g., Peixoto et al., 2017). Thus, positive findings should be interpreted considering these low scores.

In addition, as expected, the small benefits found for the retention group are short-lived since our long-term results did not reveal differences between retained and promoted students on the psychosocial variables assessed as students progressed to upper secondary education. This finding, despite being only exploratory, is in line with previous studies investigating the long-term effects of grade retention on students' self-concept, self-esteem, and goal orientations, revealing, once more, that the positive effects of grade retention tend to vanish along the way (Goos et al., 2021; Hwang & Cappella, 2018; Jimerson, 2001; Martin, 2011; Wu et al., 2010).

Finally, contrary to our predictions, the results did not demonstrate that grade retention increases the chances of subsequent grade retention, as found in previous studies (Geng & Rockoff, 2017; Mendes et al., 2015). It should be mentioned, however, that these results refer to students' upper secondary education, which has a different organisation and different grade retention standards and regulations. Based on this, we cannot advocate that grade retention prevents students' from being further retained. Moreover, the low participation rate in this longitudinal study may provide useful insights regarding students' school career. We hypothesise that, since grade retention is associated with an increase in students' mobility to different schools or alternative school tracks (Geng & Rockoff, 2017; Goos et al., 2013; Mendez et al., 2015), it might be the case that students that were further retained changed schools between Y3 and W4.

Limitations and Future Directions

Despite the contributions of this study, our findings must be interpreted in light of its limitations. First, our sample is considerably small, and thus, our findings should be generalised with caution. Also, the small sample size may have prevented us from reaching statistical significance in some results due to weak test power. Studies investigating grade retention often deal with sample size issues (e.g., Klapproth et al., 2016; Kretschmann et al., 2019) because they focus on a specific group of students, which corresponds to a smaller

group of the school population. This issue is more relevant for longitudinal studies, given their increased odds of sample mortality and missing responses.

Second, some remarks regarding our measurement waves should be made. We collected information in the middle of the school year, which gives some advantages for pre-treatment data (Kretschmann et al., 2019). However, we could not estimate the immediate short-term effects of grade retention, as in the study of Mathys et al. (2019), which collected data at the beginning of the retention year. In addition, we have different pre-retention and post-retention measurement points for the different groups of retained students, meaning that for 7th grade repeaters, we have one pre-retention measurement point and two post-retention measurement points, and for 8th grade repeaters, we have two pre-retention measurement points and only one post-retention measurement point. Finally, our long-term data has a time gap, which has prevented us from investigating grade retention effects in a unique model. Future studies should consider these design challenges to provide equal and continuous measurement points for each retention group.

Third, in our study, we did our best to employ matching methods, as was highly recommended in previous reviews (Allen et al., 2009; Goos et al., 2021; Valbuena et al., 2021). However, we should acknowledge that the limited number of baseline variables and covariates collected hindered the matching of promoted and retained students and contributed to sample reduction. Collecting important information related to grade retention could be challenging for studies using original and longitudinal data, given the volume of the data needed (Goos et al., 2021; Moser et al., 2012), but also given the sensitivity of the data (e.g., students' household situation) to be collected. Nevertheless, we recognise that this method is preferable over an outcome analysis controlling for covariates (Austin & Stuart, 2015). Thus, future studies should focus on collecting comprehensive information about students' characteristics and contexts to increase the quality of the matching procedures.

A fourth limitation should be mentioned. We failed to employ multilevel models in our data analysis due to the small number of observations for each school. Since we are investigating a phenomenon within a school context, multilevel models would be highly advisable (Marsh, 2016). In addition, we could not investigate the effects of the school context, i.e., the proportion of repeaters in school, on the relationship between grade retention and psychosocial outcomes. Previous studies showed that school retention composition also affects students' psychosocial outcomes (Demanet & Van Houtte, 2016; Hong & Yu, 2008; Pipa & Peixoto, 2022; Van Canegem et al., 2021); thus, we recommend future studies to consider this moderating variable when assessing grade retention effects.

Fifth and final, the findings of this study are explained by students' feelings of having failed, the stigma associated with grade retention, and the increasing devaluation of academic-related tasks and activities. However, to provide empirical support for these conceptions, we

advise future research to consider other psychosocial variables, such as students' peer relationships (e.g., Demanet & Van Houtte, 2016; Pagani et al., 2001). In addition, we suggest considering collecting qualitative data through interviews with retained students to collect vivid information regarding their views and experiences of grade retention (Anderson et al., 2005; Santana, 2019).

Conclusion

Based on the findings of this longitudinal study, which generally found negligible effects of grade retention, we do not support grade retention. As some authors have previously stated, given the economic, time, and societal costs of grade retention, studies must consistently find positive effects of this practice to consider it a potentially beneficial intervention for students (Allen et al., 2009; Jimerson, 2001; Goos et al., 2021; Valbuena et al., 2021). This was not the case in this study, nor the case of most high-quality studies, which tend to find null effects of grade retention (Allen et al., 2009; Goos et al., 2021; Valbuena et al., 2021). In addition, we can only recommend promoting students to the next grade level with additional intervention since at-risk promoted students also scored considerably low on psychosocial outcomes. Instead, we advocate for educators and policymakers to consider other interventions and practices for struggling students. A wide range of interventions and practices, such as parental engagement, parental engagement, reading interventions, increasing instructional time, peer or 1-to-1 tutoring, summer schools, or curriculum flexibility, are cost-effective in acting or preventing students' learning difficulties (Higgins et al., 2022; Valbuena et al., 2021). In addition, and based on the findings of this study, interventions that consider students' self-concept and motivation should also be considered to prevent students from experiencing grade retention (Higgins et al., 2022; Smith et al., 2022).

References

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A Meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis, 31*, 480–499. https://doi.org/10.3102/0162373709352239

Almeida, L. S., & Lemos, G. C. (2007). Bateria de provas de raciocínio: Manual técnico [Battery of reasoning test: Technical manual]. Universidade do Minho.

Anderman, C., & Johnson, D. (2002). The development of goal orientation. In A. Wigfield, & J. S. Eccles (Eds.), *Educational Psychology: Development of Achievement Motivation*, pp. 197-220. Academic Press. https://doi.org/10.1016/B978-012750053-9/50010-3.

Anderson, G.E., Jimerson, S.R., & Whipple, A.D. (2005) Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology, 21*, 1–20. https://doi.org/10.1300/J370v21n01_01

Austin, P. C. (2011). An introduction to propensity score methods for reducing confounding effects in observational studies. *Multivariate Behavioral Research, 46*, 399–424. https://doi.org/10.1080/0027 3171.2011.568786

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine, 34*, 3661–3679. https://doi.org/10.1002/sim.6607

Belot, M., & Vandenberghe, V. (2014). Evaluating the "threat" effects of grade repetition: Exploiting the 2001 reform by the French-speaking community of Belgium. *Education Economics, 22*, 73-89. https://doi.org/10.1080/09645292.2011.607266

Conselho Nacional de Educação (2015). *Retenção escolar nos ensinos básico e secundário* [Grade retention in primary and secondary education]. Conselho Nacional de Educação. https://www.cnedu.pt/pt/publicacoes/outros/1011-relatorio-tecnico-retencao-nos-ensinos-basico-e-secundario

Conselho Nacional de Educação (2020). *Estado da educação 2019* [State of Education 2019]. Conselho Nacional de Educação. https://www.cnedu.pt/pt/publicacoes/estado-da-educacao/1615-estado-da-educacao-2020

Decree-Law No. 55/2018 (2018). Presidência do Conselho de Ministros. Diário da República – Série I – No. 129 de 6 de Julho, pp. 2928-2943. https://data.dre.pt/eli/dec-lei/55/2018/07/06/p/dre/pt/

Demanet, J., & Van Houtte, M. (2016). Are flunkers social outcasts? A multilevel study of

grade retention effects on same-grade friendships. *American Educational Research Journal, 53*, 745– 780. https://doi.org/10.3102/0002831216646867

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review, 52*, 245–273. https://doi.org/10.1086/528764

Ehmke, T., Drechsel, B., & Carstensen, C. H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. *Studies in Educational Evaluation, 36*, 27–35. https://doi.org/10.1016/j.stueduc.2010.10.003

Ehmke, T., Salzer, C., Pietsch, M., Drechsel, B., & Muller, K. (2017). Competence development in the school year after PISA 2012: Effects of grade retention. *Zeitschrift fur Erziehungswissenschaft, 20*, 99–124. https://doi.org/10.1007/s11618-017-0752-4

Eurydice. (2022). *Key features of the education system: Portugal*. https://eurydice.eacea.ec.europa.eu/national-education-systems/portugal/portugal

European Commission. (2020). *Equity in school education in Europe. Structures, policies and student performance*. Eurydice report. Luxembourg: Publications Office of the European Union. https://doi.org/10.2797/286306

Geng, T., & Rockoff, J. (2017*). Does repeating a grade make students (and parents) happier? Regression discontinuity evidence from New York City*. https://www.udesa.edu.ar/sites/default/files/rockoff.pdf

Giano, Z., Williams, A. L., & Becnel, J. N. (2022). Grade retention and school dropout: Comparing specific grade levels across childhood and early adolescence. *Journal of Early Adolescence, 42*, 33–57. https://doi.org/10.1177/02724316211010332

Goos, M., Pipa, J., & Peixoto, F. (2021). Effectiveness of grade retention: A systematic review and meta-analysis. *Educational Research Review, 34*. 100401. https://doi.org/10.1016/j.edurev.2021.100401

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*, 323–347. https://doi.org/10.1016/j.jsp.2013.03.002

Gubbels, J., van der Put, C. E., & Assink, M. (2019). Risk factors for school absenteeism and dropout: A meta-analytic review. *Journal of Youth and Adolescence 48*, 1637–1667. https://doi.org/10.1007/s10964-019-01072-5

Higgins, S., Katsipataki, M., Villanueva Aguilera, A. B., Dobson, E., Gascoine, L., Rajab, T., Kalambouka, A., Reardon, J., Stafford, J., & Uwimpuhwe, G. (2022). The Teaching and Learning Toolkit: Communicating research evidence to inform decision-making for

policy and practice in education. *Review of Education, 10*, e3327. https://doi.org/10.1002/rev3.3327

Hong, G. L., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44,* 407–421. https://doi.org/10.1037/0012-1649.44.2.407

Hwang, S. H., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness, 11*, 559-587. https://doi.org/10.1080/19345747.2018.1496500

Jimerson, S. R. (2001). A synthesis of grade retention research: Looking backward and moving forward. *California School Psychologist, 6*, 47–59. https://doi.org/10.1007/BF03340883

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22*, 314-339. https://doi.org/10.1037/1045-3830.22.3.314.

Kaplan, A., & Maehr, M. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review, 19*, 141-184. https://doi.org/10.1007/s10648-006-9012-5

Klapproth, F., & Schaltz, P. (2015). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. *European Journal of Psychology of Education, 30*, 119–136. https://doi.org/10.1007/s10212- 014-0232-7

Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences, 50*, 182-194. https://doi.org/10.1016/j.lindif.2016.08.014

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology, 111*, 1432–1446. https://doi.org/10.1037/edu0000353

Lamote, C., Pinxten, M., Van Den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies, 40*, 496–514. http://doi .org/10.1080/03055698.2014.936828

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative

effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology, 108*, 256–273. https://doi.org/10.1037/edu0000059

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109*, 425. doi:10.1037/edu0000144

Marsh, H. W., Trautwein, U., Lüdke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*, 510-524. https://doi.org/10.1037/0022-0663.100.3.510

Martin, A.J. (2009). Age appropriateness and motivation, engagement, and performance in high school: Effects of age within cohort, grade retention, and delayed school entry. *Journal of Educational Psychology, 101*, 101-114. https://doi.org/10.1037/a0013100

Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal, 37*, 739–763. https://doi.org/10.1080/01411926.2010.490874

Mathys, C., Véronneau, M.-H., & Lecocq, A. (2019). Grade retention at the transition to secondary school: Using propensity score matching to identify consequences on psychosocial adjustment. *Journal of Early Adolescence, 39*, 97–133. doi:10.1177/0272431617735651

Mendez, L. M. R., Kim, E. S., Ferron, J., & Woods, B. (2015). Altering school progression through delayed entry or kindergarten retention: Propensity score analysis of long-term outcomes. *Journal of Educational Research, 108*, 186-203. https://doi.org/10.1080/00220671.2013.867474

Ministério da Educação/Direcção Geral de Educação. (n.d.). Estatísticas do ensino básico e secundário. https://infoescolas.medu.pt

Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology, 104*, 603–621. https://doi.org/10.1037/a0027571

Nascimento, S., & Peixoto, F. (2012). Relações entre o estatuto escolar e o autoconceito, auto-estima e orientações motivacionais em alunos do 9º ano de escolaridade [Relationship between school status, self-esteem and goal orientations among 9th grade students]. *Análise Psicológica, 30*, 421-434. https://doi.org/10.14417/ap.602.

Nunes, L. C., Balcão Reis, A., & Seabra, C. (2018). Is retention beneficial to low-achieving students? Evidence from Portugal. *Applied Economics, 50*, 4306–4317. https://doi.org/10.1080/00036846.2018.1444261

Organisation for Economic Co-operation and Development (2020). *PISA 2018 results: Effective policies, successful schools* (Vol. 5). OECD Publishing. https://doi.org/10.1787/ca768d40-en

Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology*, 13, 297–315. https://doi.org/10.1017/s0954579401002061

Peixoto, F., & Almeida, L. S. (2010). Self-concept, self-esteem and academic achievement: Strategies for maintaining self-esteem in students experiencing academic failure. *European Journal of Psychology of Education, 25,* 157–175. https://doi.org/10.1007/s10212-010- 0011-z

Peixoto, F., Monteiro, V., Mata, L., Sanches, C., Pipa, J., & Almeida, L. (2016). "To be or not to be retained…That's the question!" Retention, self-esteem, self-concept, achievement goals, and grades. *Frontiers in Psychology, 7*, 1550. https://doi.org/10.3389/fpsyg.2016.01550

Peixoto, F., Pipa, J., Mata, L., Monteiro, V., & Sanches, C. (2017). "To learn, or to be the best?" Achievement goal profiles in pre-adolescents. *Análise Psicológica, 35*, 499-511. https://doi.org/10.14417/ap.1182

Pereira, M. C., & Reis, H. J. (2014). Grade retention during basic education in Portugal: Determinants and impact on student achievement. *Economic Bulletin, 16,* 61–83. https://www.bportugal.pt/sites/default/files/anexos/papers/ab201407_e.pdf

Pipa, J., & Peixoto, F. (2022). One step back or one step forward? Effects of grade retention and school retention composition on Portuguese students' psychosocial outcomes using PISA 2018 data. *Sustainability, 14*, 16573. https://doi.org/10.3390/ su142416573

Pipa, J., Peixoto, F., Mata, L., Monteiro, V., & Sanches (2016). The Goal Orientations Scale (GOS): Validation for Portuguese students. *European Journal of Developmental Psychology, 14*, 477-488. doi:10.1080/17405629.2016.1216835

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. https://doi.org/10.1002/9780470316696

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46,* 407–441. https://doi.org/10.2307/1170010

Skaalvik, E. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology, 89*, 71-81. https://doi.org/10.1037/0022-0663.89.1.71

Santos, N., Pipa, J., & Monteiro, V. (2023). Analysing grade retention beliefs within teachers'

psycho-pedagogic beliefs system. *Teaching and Teacher Education*, *121.* https://doi.org/10.1016/j.tate.2022.103939

Smith, J., Paquin, S., St-Amand, J., Singh, C., Moreau, D., Bergeron, J., & Leroux, M. (2022). A remediation measure as an alternative to grade retention: A study on achievement motivation. *Psychology in the Schools, 59*, 1209-1221. https://doi.org/10.1002/pits.22677

Smith, M. L., & Shepard, L. A. (1988). Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices. *American Educational Research Journal, 25*, 307–333. https://doi.org/10.2307/1163167

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science, 25*, 1–21 https://doi.org/10.1214/09-STS313

Valbuena, J., Mediavilla, M., Choi, A., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys, 35*, 408-451. https://doi.org/10.1111/joes.12406

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software, 45*, 1-67. https://doi.org/10.18637/jss.v045.i03

Van Canegem, T., Van Houtte, M., & Demanet, J. (2021). Grade retention and academic self-concept: A multilevel analysis of the effects of schools' retention composition. *British Educational Research Journal, 47*, 1340-1360. https://doi.org/10.1002/berj.3729

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). Time-varying treatments in observational studies: Marginal structural models of the effects of early grade retention on math achievement. *Multivariate Behavioral Research, 51*, 843-864. https://doi.org/10.1080/00273171.2016.1155146

van der Wal, W. M., & Geskus, R. B. (2011). ipw: an R package for inverse probability weighting. *Journal of Statistical Software, 43*, 1-23. https://doi.org/10.18637/jss.v043.i13

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology, 102*, 135–152. https://doi.org/10.1037/a0016664

Xiang, N., & Chiu, S. (2022). The school matters: Hong Kong secondary schools' grade-retention composition, students' educational performance, and educational inequality. *School Effectiveness and School Improvement*. https://doi.org/10.1080/09243453.2022.2136210

SECTION III. GENERAL DISCUSSION

6. General Discussion

In this concluding Chapter, we summarise and discuss the main findings of this dissertation, bringing together the results of the empirical studies. After, we elaborate on the limitations of this work and provide suggestions for future research. Finally, implications for educational policy and practice are addressed.

6.1. Main Research Findings

In this dissertation, we investigated the effects of grade retention in Portuguese lower secondary education on the affective components of learning. Our main goals were to provide empirical evidence on 1) the effectiveness of grade retention through a meta-analysis (Study 1, chapter 3) estimating the effects of grade retention on several academic and non-academic outcomes and considering several moderating variables; and 2) the effects of grade retention within the Portuguese context on the affective components of learning, through a retrospective cross-sectional study using a representative sample (Study 2, chapter 4), and a longitudinal study providing evidence on the short-medium-and long-term effects of grade retention (Study 3, chapter 5).

Study 1 (chapter 3) reviewed 84 studies investigating grade retention's academic and non-academic effects. Our meta-analysis added to the previous reviews (Allen et al., 2009; Bright, 2011; Holmes, 1989; Holmes & Matthews, 1984; Jackson, 1975; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009) in that we integrate studies investigating primary and secondary grade retention; considering a variety of countries across the world; and using sound methods to estimate grade retention effects. In addition, we introduced several moderators to investigate whether grade retention effects differed according to several study, sample, and context characteristics. In line with Allen et al.'s (2009) review, we found an overall 'zero' effect of grade retention. However, and most importantly, our review showed that grade retention seems to be more effective: a) in Education systems such as the United States that often couples grade retention with other interventions and support; b) considering psychosocial variables such as students' academic behaviour, self-concept, etc.; c) in short-term; d) studies using same-grade comparison. On the other hand, grade retention seems least effective a) in countries with tracking systems, such as Belgium, Germany, or Luxembourg; b) in medium and long-term; and c) in studies using same-age comparison. Additionally, although not statistically significant, we found an overall negative effect size in studies considering Education systems with a uniform organisation, where Portugal is included (Dupriez et al., 2008; Mons, 2007). Finally, studies falling into the category of psychosocial factors presented a great variety of psychosocial outcomes, from studies considering affective components and learning, such as self-concept, and motivational aspects (e.g., Ehmke et al., 2010, 2017;

Kretschmann et al., 2019; Mathys et al., 2019; Wu et al., 2010), to those considering juvenile crime and other behavioural conduct (e.g., Eren et al., 2017; Özek, 2015).

These latter two considerations inspired the remaining studies of this dissertation. Remarkably, little is known about the effects of grade retention on students' sense of school belonging as a measure of students' sense of identification with school (Finn, 1989; Voelkl, 1996), as well as the effects of grade retention on students' goal orientations (Peixoto et al., 2016, 2017). Thus, in chapter 4, we presented a cross-sectional study using representative data of the Portuguese context – PISA 2018 – that estimated the effects of grade retention in grades 1 to 9 on 15-year-old students' reading self-concept, goal orientations, school belonging, and school valuing. The results showed detrimental effects on students' task orientation and sense of school belonging and valuing, findings that are in line with previous studies (Klapproth et al., 2016; Martin, 2011; OECD, 2019; Peixoto et al., 2016, 2017; Van Canegem et al., 2021, 2022b). After retention, students are detached from their peer group, which could trigger feelings of not being accepted, respected, or valued within their school community. Thus, from a social goal perspective (Wentzel, 2002) and the stage-environment fit framework (Eccles &b Midgley, 1989; Eccles et al., 1993), students would feel disengaged in contexts that fail to provide opportunities to establish and maintain personal relationships and when they feel a misfit between their personal needs and the educational environment offered by the school. Most notably, we found that in schools with a larger proportion of retained students, the whole student community showed lower self-concept in reading, task orientation and sense of school valuing. Moreover, promoted students are those more affected by the number of repeaters in school. These results stress that the adverse effects of grade retention extend to their classmates and the broader school community (Bietenbeck, 2014; Demanet & Van Houtte, 2013, 2016; Gottfried, 2013; Van Canegem et al., 2021). Nevertheless, this study's cross-sectional nature prevented us from drawing causal inferences regarding the effects of grade retention on these components and, what is more, disentangling the short vs long-term effects of retention.

In Study 3 (chapter 5), we addressed this limitation by conducting a longitudinal study investigating the effects of grade retention in lower secondary education on the trajectories of students' academic self-concept, self-esteem, and goal orientations. In addition, we explored the long-term effects of retention on these components and students' school career. Our findings did not mirror the idea that grade retention would motivate students to work harder (Belot & Vandenberghe, 2014; Dupriez et al., 2008; Hong & Raudenbush, 2005; Range et al., 2012), given the lower levels of self-concept and goal orientations observed before retention in the groups of retained students (Kretschmann et al., 2019). Moreover, except for positive short-term effects found in academic self-concept (Hwang & Cappella, 2018; Jimerson, 2001; Marsh, 2016; Marsh et al., 2017; Marsh, Seaton, et al., 2008; Marsh, Trautwein, et al., 2008;

Martin, 2011; Wu et al., 2010), no significant differences were found between retained students across the years, suggesting that repeaters seemingly would have developed (at least) similar self-esteem and motivational orientations, if promoted instead of being retained (Hwang & Cappella, 2018; Klapproth et al., 2016; Kretschmann et al., 2019; Lamote et al., 2014; Martin, 2011). The findings regarding students' avoidance orientations also suggest that retained students may continuously devalue academic-related tasks to avoid failure and protect their feelings of self-worth (Covington, 2000; Nascimento & Peixoto, 2012; Peixoto et al., 2016, 2017). Finally, our exploratory results did not reveal differences between retained and promoted students in subsequent grade retention. However, we hypothesised that this could also be due to effects on students' school career via increased school mobility and enrolment in different school tracks (Geng & Rockoff, 2017; Goos, Van Damme, et al., 2013; Moser et al., 2012; Mendez et al., 2015).

Altogether, the results from the three empirical studies denote an overall null or negative effect of grade retention. In Study 1, we found an overall zero effect of grade retention across several academic and non-academic outcomes. Additionally, Studies 2 and 3 addressing specific affective components of learning, roughly found negative to nonsignificant effects of grade retention on students' goal orientations and school belonging and valuing. Thus, these findings do not support grade retention nor consider it an effective practice because, as claimed by several authors, a null effect regarding a measure that entails such high costs calls into question its effectiveness and its educational benefits (Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2021; Xia & Kirby, 2009).

Several arguments are often raised to defend the application of retention (for an overview see chapter 1.1.). One of those arguments is that students would engage in more positive learning behaviours and work harder if they knew that grade retention could be a possible consequence of their academic failure (Dupriez et al., 2008; Belot & Vanderberghe, 2014; Hong & Raudenbush, 2005; Range et al., 2012). This argument emerges from teachers' beliefs that grade retention acts in students' best interests and, thus, applied it as both a corrective and pedagogical measure (Alet et al., 2013; Santos et al., 2023). However, as the results from Study 3 indicated, students' self-concept and goal orientations are already 'at risk' before being retained, as was also found in Kretschmann et al.'s (2019) study, which contradicts the idea that the threat of being retained would motivate students to academic performance. These results add to the few existing studies that consider motivational and other psychosocial variables as predictors of grade retention, such as students' self-concept (Ferrão, 2015), peer acceptance (Jimerson et al., 1997; Lubbers et al., 2006; Sidali et al., 2021), motivational orientations (Gadeyne et al., 2008; Kretschmann et al., 2019; Sidali et al., 2021), and emotional health (Alexander et al., 2003; Gadeyne et al., 2008; Huang, 2014; Jimerson et

al., 1997), calling into attention for the need of monitoring and intervening on these aspects preventively before retention occurs (Smith et al., 2022).

Regarding grade retention effects, we found that being retained could have potential short-term gains, as shown in Study 1 and Study 3. In Study 1, we found positive short-term effects across all outcome variables. This finding was also found in Study 3, where 7th grade repeaters increased their self-perception of academic competence after being retained. Positive effects of grade retention on students' academic self-concept were also found in other studies (Ehmke et al., 2010; Lamote et al., 2014; Marsh et al., 2017), and this finding is commonly explained by social comparison theory (Festinger, 1954), and the big-fish-little-pond effect (Marsh, 2016; Marsh et al., 2017; Marsh, Seaton, et al., 2008; Marsh, Trautwein, et al., 2008). In this framework, it is acknowledged that by changing their comparison reference group to their younger and less experienced classmates, retained students may perceive themselves as more competent academically. In addition, by repeating the same grade, repeaters already know the rules, which offers them some advantage over their colleagues, and may finally experience success and gain positive feedback from their teachers, which also could lead to more positive feelings of academic competence (Ehmke et al., 2017; Gleason et al., 2007; Goos, Van Damme, et al., 2013; Smith et al., 2022).

On the other hand, these short-term gains were not mirrored in goal orientations. What is more, Studies 1, 2 (potentially), and 3 add to the existing literature by showing that the positive effects of grade retention are very short-lived (Allen et al., 2009; Hwang & Cappella, 2018; Jimerson, 2001; Martin, 2011; Moser et al., 2012; Valbuena et al., 2021; Wu et al., 2010). Study 1 found that grade retention effects tend to decrease over time. Regarding Study 2, we can only speak of potential medium to long-term detrimental effects given that we could not estimate with precision in grade students were retained with precision. Nevertheless, we found detrimental effects of grade retention on students' task orientation, school belonging and valuing. Study 3 mainly found medium and long-term nonsignificant effects in goal orientations and academic self-concept. Together, these findings stress the idea that grade retention brings about heavy emotions and feelings of having failed that prevail over time (Alexander et al., 2003; Anderson et al., 2005; Yamamoto & Byrnes, 1987; Jimerson & Ferguson, 2007; Santana, 2019) and these feelings, in turn, affect students' motivation, engagement and self-esteem (Goos, Van Damme, et al., 2013; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Van Canegem et al., 2021; Xiang & Chiu, 2022).

Additionally, the lower sense school belonging, and school valuing demonstrated by retained students in Study 2 and the low students' importance attributed to academic self-concept and goal orientations found in Study 3 suggest that retained students may experience a mismatch between their needs and the school environment, as posited by the stage-environment fit theory (Eccles & Midgley, 1989; Eccles et al., 1993), particularly considering

secondary education. In this sense, the mere repetition of school subjects resulting from grade retention (Pagani et al., 2001; Bonvin et al., 2008) may undermine students' motivation due to deprivation of a new and challenging environment (Goos, Van Damme, et al., 2013; Hong & Yu, 2008). Also, retained students lost their peer group who got promoted to the next grade level, which may trigger both negative self-feelings and lower acceptance by their new classmates (Goos et al., 2013; Demanet & Van Houtte, 2016; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Pagani et al., 2001; Van Canegem et al., 2021; Wu et al., 2010; Xiang & Chiu, 2022). Thus, based on these interconnected experiences, retained students may be more prone to feel that the school environment does not correspond to their needs, which in turn, leads to lower motivation and engagement at school and lower feelings of school belonging (Demanet & Van Houtte, 2013, 2016; Eccles et al., 1993; Finn, 1989; Martin, 2011; Pagani et al., 2001; Van Canegem et al., 2021, 2022; Wu et al., 2010).

The abovementioned findings considering students' affective components of learning would be better interpreted by keeping students' achievement in sight. In fact, we did not include academic achievement in our analyses, given the goals of the studies, yet, we acknowledge that these two components act together in understanding the effects of grade retention. Results regarding the development of students' achievement in Math and Portuguese subjects from Study 3 are presented in Appendix C. Our results are in line with previous studies by showing that small positive effects in achievement tend to vanish along the way (e.g., Allen et al., 2009; Jimerson 2001; Klaproth et al., 2016; Moser et al., 2012; Valbuena et al., 2021; Wu et al., 2008). In addition, it was found that grade retention was negatively associated with students' performance in PISA study (OECD, 2020). Thus, we could conclude that grade retention did not improve students' affective components of learning nor their academic achievement.

Self-worth theory (Covington, 1992; 2000) provide valuable insights to better understand the results of this study. This theory posits that the search for self-acceptance is the highest human priority (Castella et al., 2013). Since school experiences play a central role in their lives, school-aged children and youth naturally seek acceptance in this context. Acceptance and sense of worth in the school context result from students' achievement and academic competence, and thus, performing poorly may lead to feelings of shame and humiliation (Castella et al., 2013; Covington, 2000). As a consequence, struggling students, here represented as repeaters, as a way of protecting their feelings of self-worth, would engage in self-handicapping and helplessness behaviours by decreasing their expectations and by disinvesting or avoiding school-related tasks (Castella et al., 2013; Covington, 1992; 2000; Nascimento & Peixoto, 2012; Peixoto & Almeida, 2010; Peixoto et al., 2016, 2017). In this dissertation, this is visible in the stable and relatively higher levels of self-esteem and the upward trajectory in avoidance orientation presented in Study 3.

Nevertheless, it should be mentioned that exposure to repeated failure, whether it would be considered as subsequent retention (e.g., Geng & Rockoff, 2017; Mendes et al., 2015), changes in school career (Geng & Rockoff, 2017; Goos, Van Damme, et al., 2013; Moser et al., 2012; Mendez et al., 2015) or just struggling academically, may push students to continually create more unrealistic excuses, which in turn will lose their self-protective value (Covington, 2000). It is also noteworthy that grade retention has been found to be strongly associated with the following long-term adverse outcomes: school dropout, lower postsecondary education enrolment, and less favourable professional career (Allensworth, 2005; Fine & Davis, 2003; Giano et al., 2022; Gubbels et al., 2019; Fraysier et al., 2020; Jimerson, 1999; Jimerson et al., 2002; Ou & Reynolds, 2010; Santos et al., 2022; Schwedt et al., 2017). This consideration, combined with the findings of this work, suggests that grade retention harms students' academic development.

At this point, it should be stressed that the present study's findings also revealed some weaknesses of social promotion, particularly in Study 3, where promoted students also presented low self-concept, self-esteem, and goal orientations. This result, together with the findings from 'retention plus' moderating effects from Study 1 and the impact of school retention composition on promoted students' outcomes from Study 3, challenges the view that social promotion *alone* would be an effective alternative for retention. Instead, based on these findings, we argue that support must be given to *all* struggling students over and beyond their retention history (Conboy et al., 2013; Goos, Van Damme, et al., 2013; Valbuena et al., 2021).

This dissertation also aimed to contribute to the research findings regarding the moderating effects of school retention composition (Bietenbeck, 2014; Demanet & Van Houtte, 2013, 2016; Gottfried, 2013; Santos et al., 2022; Van Canegem et al., 2021; Van Canegem et al., 2022a, 2022b). We found that school retention composition negatively predicted students' reading self-concept, task orientation and students' school valuing (see Study 2, Chapter 4). In addition, we found a moderating effect of school retention composition on the relationship between retention and students' self-concept, task orientation, school belonging and school valuing. These findings add to the previous research that supports the moderating effects of retention composition on students' academic achievement (Bietenbeck, 2014; Gottfried, 2013; Hong & Raudenbush, 2005), same-grade friendships (Damenet & Van Houtte, 2016), school misconduct and academic behaviour (Bietenbeck, 2014; Demanet & Van Houtte, 2013), school belonging (Van Canegem et al., 2021), and postsecondary enrollment intentions (Santos et al., 2022).

Studying this moderating effect not only reflects the hierarchical nature of the data (Bastos & Ferrão, 2019; Marsh, 2016) but also, and most importantly, provides important insights regarding the contextual effects of grade retention. Based on our findings, grade retention exerts a spillover effect over their promoted peers (Bietenbeck, 2014; Gottfried,

2013), questioning, once more, the effectiveness of this practice, which also impacts the school environment and non-retained students. Indeed, having many repeaters in school could disrupt the environment, undermining students' engagement (Demanet & Van Houtte, 2016; Gottfried, 2013; Lavy et al., 2012; Smith & Shepard, 1988). The fact that the proportion of repeaters in school affects mainly non-retained students could be explained given that being a repeater in a school with several others in the same condition reduces the often-mentioned stigmatising effect of retention (Demanet & Van Houtte, 2013; Pagani et al., 2001; Wu et al., 2010). Moreover, when the practice of retention is widespread among schools and countries (Agasisti & Cordero, 2017; Goos, Schreier, et al., 2013; Van Canegem et al., 2022a, 2022b), it brings some normalisation to be a repeater, which could explain nonsignificant effects on students' psychosocial outcomes (Klapproth et al., 2016). Indeed, in Van Canegem et al. (2022a, 2022b) studies using PISA data, country retention rates moderated the relationship between grade retention and students' belonging and bullying victimisation. These studies suggest that retained students stand out in contexts where grade retention is solely applied and, thus, grade retention becomes a more negative experience in terms of students' psychosocial outcomes when there are few retained students within the education system. The same findings were presented by OECD (2020) regarding PISA performance scores. Together, these findings could also explain the less detrimental effects on some affective components of learning obtained in this dissertation, given the fact that in Portugal, grade retention is still often applied at both school and country levels and may, thus, be seen as a 'natural and expected endpoint' for students.

Finally, concerning these latter findings, we must mention that the proportion of retained students in a school is closely linked with other socioeconomic and cultural features of that school, such as the share of disadvantaged students, students with an immigrant background and low achievement (Bastos e Ferrão, 2019; Lavy et al., 2012). Furthermore, as discussed in Chapter 2.2, we know these characteristics are highly associated with a greater retention risk. Thus, the widespread of grade retention in a context with such unequal social and economic conditions would only increase the disparities between students, exacerbate inequalities among students, and undermine education equity, as was also pointed out by OECD (2020). These results challenge the view of the effectiveness of grade retention at the meso (school) and macro levels (country).

6.2. Limitations and directions for future studies

With the present dissertation, we aimed to contribute to the existing body of knowledge regarding the effects of grade retention on the affective components of learning. Particularly, we intended to add to the existing studies in the Portuguese context, especially given the lack of studies employing high-quality methods and investigating the effects of grade retention on

psychosocial outcomes in this context. Nevertheless, each study presented in this work has its strengths and limitations that were already addressed in the previous chapters. In this section, we acknowledge some general limitations of this dissertation.

First, considering the findings of grade retention effects on psychosocial variables obtained in our meta-analysis (Study 1), we failed to further explore which psychosocial variable (or group of variables) grade retention would produce more or less positive effects. Indeed, the effect sizes found ranged between -2.58 and 1, showing that grade retention could be detrimental for some psychosocial aspects. In addition, as shown in Studies 2 and 3, we recognise the complexity and uniqueness of each psychosocial variable, which may be differently affected by grade retention. Hattie (2009), in his synthesis of meta-analysis on students' achievement, explored the contribution of, for example, motivational variables, self-concept variables, and engagement variables, as indicators of students' attitudes and dispositions, without mingling these variables. Thus, we suggest further investigating the impact of retention on different psychosocial variables by reviewing existing studies considering those aspects of learning to help clarify the effects of retention on these variables.

Second, except for Study 1, the remaining studies rely only on Portuguese data. We already stressed that investigating the effects of grade retention within the Portuguese context was crucial, given the limited number of studies investigating this context with such high retention rates and positive beliefs regarding the effectiveness of retention (Santana, 2019; Santos et al., 2023). However, at least for Study 2, a cross-national comparison would have been helpful to compare the findings from Portugal with other education systems. Considering an international perspective would be worthwhile to learn from other systems, especially those avoiding the application of retention (e.g., Iceland, Slovenia, Sweden, or Finland) (Agasisti & Cordero, 2017; Goos, Schreirer et al., 2013; Van Canegem et al., 2022a, 2022b).

Third, the data used in Studies 2 and 3 were observational. The data from Study 2 were cross-sectional, which prevented us from making causal claims. In addition, despite our best efforts to employ sound methodologies to estimate the causal effects of grade retention (Allen et al., 2009; Goos et al., 2021; Valbuena et al., 2021), as the data presented in Studies 2 and 3 served other purposes, several important pre-retention variables were not collected and only a limited number of covariates were considered, in comparison with other studies such as that from Goos, Van Damme, et al. (2013) and Moser et al. (2012) who used an extensive number of covariates. Thus, we could not prevent our findings from possible measurement error and unmeasured bias, and the findings of this work should be interpreted according to these limitations.

Fourth and related to the previous limitation, or selected covariates were mainly on students' individual and inalterable characteristics, such as students' gender, age, economic, social, and cultural background, and to a lesser extent, family variables (Study 2) and school

variables (Studies 2 and 3). Indeed, most studies considering the predictors of grade retention focused mainly on students' demographics (e.g., Bastos & Ferrão, 2019; Ferrão, 2015; García-Perez et al., 2014; Klapproth & Schaltz, 2015; Nunes et al., 2018; Pereira & Reis, 2014). However, family variables such as parental engagement and support have also been related to students' likelihood of grade retention and school success (Agasisti & Cordero, 2017; Almeida et al., 2005). In addition, we failed to consider also teachers' views about grade retention. We know from previous literature that teachers' beliefs regarding grade retention play a significant role in their practices (Bonvin, 2008; Santana, 2019; Santos et al., 2023; Valbuena et al., 2021). Thus, it is crucial to consider such variables as predictors of grade retention, to better contextualise and understand their impact on academic outcomes. At this point, it is noteworthy that this dissertation was conducted during a disruptive period. The COVID-19 pandemic was installed when a new empirical study designed to tackle the limitations mentioned in the last two paragraphs was ongoing. In that study, we aimed to include an extensive list of variables related to grade retention, including those considering parental engagement and teachers' beliefs regarding grade retention and their attitudes and evaluations (Bonvin et al., 2008). However, due to school closure and the lockdown situation, and anticipating that COVID-19 would impact grade retention practices (NASP, 2021) and students' motivation (Camacho et al., 2021), we have decided not to continue the study. Based on this, we encourage future research investigating the effects of grade retention within the Portuguese context to tackle these issues by collecting a comprehensive number of variables related to grade retention.

Notably, some investigation has been conducted regarding teachers' beliefs about grade retention and its relationship with the practice of grade retention (e.g., Bonvin et al., 2008; Range et al., 2012; Santana, 2019; Santos et al., 2023; Tomchin & Impara, 1992). However, little has been done so far to link grade retention beliefs, their pedagogical practices, retention practice and the effects of grade retention, particularly on students' affective components of learning. To our knowledge, the study of Santos et al. (2023) was unique in associating teachers' grade retention beliefs with their pedagogical and grade retention practices. Yet, the link between these teachers' attitudes and practices and students' outcomes remains practically unexplored (Bonvin et al., 2008). Moreover, the instruments assessing teachers' grade retention beliefs (e.g., Santos et al., 2023; Tomchin & Impara, 1992; Range et al., 2012) present several items that question teachers regarding the effects of grade retention on students' affective components of learning (e.g., self-esteem, motivation). Teachers often hold positive beliefs regarding the effects of these components because they tend to whiteness short-term effects of grade retention (Fine & Davis, 2003; Tomchin & Impara, 1992). This view does not mirror research findings, especially considering long-run effects. Given this, it would be interesting to consider teachers' beliefs and practices as predictors of retention and conduct

studies integrating teachers' variables and the effects of retention on students, especially in the long term.

Fifth, several authors often mention that the adverse effects of grade retention on students' affective components of learning may be explained by the loss of peer membership, which in turn may bring about negative self-judgments, low motivation and school engagement for retained students (Goos, Van Damme, et al., 2013; Jimerson, 2001; Kretschmann et al., 2019; Martin, 2011; Van Canegem et al., 2021; Xiang & Chiu, 2022). However, our study failed to directly assess this premise by not including measures assessing peer relationships and the social position of retained students (Demanet & Van Houtte, 2016; Gleason et al., 2007; Wentzel et al., 2020). Future studies should be conducted to further support this argument.

Sixth, and in line with the previous comment, it is widely debated that after being retained, students compare themselves with their younger and unexperienced grademates (Ehmke et al., 2010; Marsh, 2016; Marsh et al., 2017), whereas other studies claimed that students still confronted their academic situation with that of their former classmates (Goos, Van Damme, et al., 2013; Kretshcmann et al., 2019; Van Canegem et al., 2021). To our knowledge, none of the studies directly asked students about their reference group. Acknowledging that students form their self-perceptions in comparison to others, we advise future studies to consider collecting information regarding with whom repeaters compare themselves after being subject to a change in their group membership. Such information would provide valuable insights regarding the effects of retention.

Seventh, most studies investigating the impact of grade retention on students' affective components of learning and other psychosocial factors are mainly based on quantitative self-report measures. Only a few studies assessed those variables qualitatively (Anderson et al., 2005; Yamamoto & Byrnes, 1987; Yamamoto & Byrnes, 1987). Further qualitative research providing vivid information on retained and promoted students' views, histories and experiences associated with grade retention are necessary to better understand how students are affected by grade retention and promotion.

Eighth, our study considered only the average effects of retention vs promotion. Assuming that grade retention will continue to exist, despite the recommendations and the empirical evidence, it would be helpful to teachers, the broader school community and policymakers to explore under which specific circumstances grade retention would be beneficial. Specifically, such studies would help better understand who does and does not benefit from grade retention.

Finally, in Studies 2 and 3, grade retention was treated as a single intervention, not considering whether students were provided additional measures during the retention year. In fact, we could not collect such information and thus assumed that students repeated the same grade level without additional support (Conboy et al., 2013). However, as noted in Study 1,

positive effects of grade retention may be found if additional support is provided to students. Studies investigating this assumption were mainly conducted in the United States of America (e.g., Allensworth, 2005; Geng & Rockoff, 2017; Özek, 2015; Schwerdt et al., 2017). Also, as Conboy et al. (2013) claimed, the debate around grade retention and 'simple' promotion may not be as productive as debating the effects of grade retention 'plus' additional support or, even more, the effects of grade retention versus other interventions. Nevertheless, few studies compared grade retention with other interventions and programs. The limited existing studies evidenced that alternatives to retention presented better results on students' academic achievement (e.g., changes in class size, differentiated instruction, and remediation motivational measures; Barata et al., 2015; Smith et al., 2022). Based on this, we strongly advise the research community to change the 'paradigm' of grade retention effectiveness research by comparing this intervention with other possible interventions.

6.3. Implications for Educational Policy and Practice

The findings of this dissertation add to the existing body of evidence regarding the effects of grade retention on the affective components of learning and suggest that grade retention does not bring about benefits for students. Grade retention is considered a highly costly intervention, both in terms of resources, finances, and time (Brophy, 2006; Ehmke et al., 2010; European Commission, 2020; Flores et al., 2013; Higgins et al., 2022; Ikeda, 2011; Ikeda & Garcia, 2014; Justino et al., 2014; Valbuena et al., 2021), affecting students' school career and delaying their entry into the labour market (Ou & Reynolds, 2010). Given this, the benefits must outweigh the detrimental and null effects in order to be considered an effective measure, which was not the case in this dissertation, nor was the case in previous research (Allen et al., 2009; Jimerson, 2001; Valbuena et al., 2021). In addition, grade retention has been found to have critical equity implications, suggesting that its adverse effects are more pronounced for disadvantaged students (European Commission, 2020; Lavy et al., 2012; OECD, 2020), and naturally, based on these claims, OECD has advised against its use.

For policymakers, teachers, and the broader school community, we strongly recommend being aware of the risk factors associated with grade retention and avoiding subjecting those students to retention to prevent them from further detrimental effects. This would be the case of students from disadvantaged economic and social backgrounds, younger students, students with lower parental support, or students with lower self-beliefs (European Commission, 2020; Kretschmann et al., 2019; Lavy et al., 2012; OECD, 2020; Xiang & Chiu, 2022). With this, we suggest that at-risk students' characteristics, resources and dispositions should be put in perspective and fully considered before making a decision to retain them (Bonvin et al., 2008).

In addition, previous findings show that teachers generally hold positive beliefs regarding grade retention (Santana, 2019; Santos et al., 2023; Tomchin & Impara, 1992) that grade retention eases instruction by creating more homogeneous classrooms. Interestingly, recent studies provided evidence against these considerations by showing that schools with many repeaters disrupt classroom environment and instruction (Demanet & Van Houtte, 2016; Gottfried, 2013; Lavy et al., 2012; Smith & Shepard, 1988) and that grade retention does not affect only retained students but also the broader school community (Demanet et al., 2013, 2016; Santos et al., 2022; Van Canegem, 2021). These findings stress the necessity of working directly with teachers and, more precisely, with the whole school staff, providing them with scientific evidence regarding the long-lasting non-beneficial effects of grade retention, on the one hand, and supporting them in finding alternative measures to retention. As we already stated, automatic grade promotion, *per se,* would not be effective either, as observed by the low scores found in Study 3. Instead, we strongly advise policymakers, teachers, and the whole school community to consider other interventions.

Considering intervention with students, we recommend early identification of students struggling academically and providing them with targeted remedial interventions. A wide range of interventions have been proven to be cost-effective in acting on students' learning difficulties, such as parental engagement, reading interventions, increasing instructional time, peer or 1-to-1 tutoring, and summer schools (Higgins et al., 2022; Jimerson, 2001; Slavin et al., 2011; Valbuena et al., 2021). In addition and grounded on the findings of this study, we advocate also considering acting preventively on students' self-concept, motivation, and engagement (Higgins et al., 2022; Smith et al., 2022).

In the case of retention, we recommend providing students with additional support rather than merely a rehearsal of the learning material. At the macro level, policymakers should consider making educational systems as curriculums more flexible (Valbuena et al., 2021), for example, by allowing students to repeat only the subjects they have failed while progressing the remaining subjects along with their peers. This measure could prevent students from experiencing low motivation, disengagement, and isolation due to retention. However, from the different retention rates observed between Education systems and considering the positive teachers' beliefs regarding retention, irrespective of its effects and legislations, we cannot assure that directly changing policies would be effective enough. We, as Valbuena et al. (2021), acknowledge that, most likely, the force of change lies on parents, teachers, and school community beliefs and practices and thus changing these beliefs, supporting those actors in improving their teaching quality and in implementing alternative practices would be, most likely, more effective.

6.4. Conclusion

Grade retention research has improved tremendously during the last two decades. With this dissertation, we aimed to contribute to this research field with high-quality results regarding the effectiveness of grade retention and the investigation of its effects on students' affective components of learning, considering a context where high retention rates and high positive beliefs regarding grade retention effectiveness still prevail. In sum, this work adds to previous research that calls into question the effectiveness of this practice, claiming that it is more harmful than helpful for students. Teachers, of course, may retain students with the best intentions. However, they probably do not have the opportunity to see its long-run detrimental effects. In the reviewed studies, grade retention was consistently more detrimental in long-term studies and, on the other hand, more effective when coupled with additional support. In addition, grade retention was associated with a lower academic self-concept, lower task orientation, and lower sense of belonging and valuing. This study also highlights the crucial role of school retention composition in understanding grade retention effects. With these findings, we call attention to the spillover effects of grade retention as it affects the broader student community. All in all, we advise against the use of grade retention or, at least, caution in its application.

References

Agasisti, T., & Cordero, J. M. (2017). The determinants of repetition rates in Europe: Early skills or subsequent parents' help? Journal of Policy Modeling, *39*, 129-146. https://doi.org/10.1016/j.jpolmod.2016.07.002

Alet, É., Bonnal, L., & Favard, P. (2013). Repetition: Medicine for a short-run remission. *Annals of Economics and Statistics, 111,* 227-250. https://doi.org/10.2307/23646332

Alexander, K. L., Entwisle, D. R., & Dauber, S. L. (2003). *On the success of failure: A reassessment of the effects of retention in the primary school grades*. Cambridge University Press.

Allen, C. S., Chen, Q., Willson, V. L., & Hughes, J. N. (2009). Quality of research design moderates effects of grade retention on achievement: A Meta-analytic, multi-level analysis. *Educational Evaluation and Policy Analysis, 31*, 480–499. https://doi.org/10.3102/0162373709352239

Allensworth, E. M. (2005). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Educational Evaluation and Policy Analysis, 27*(4), 341-364. https://doi.org/10.3102/01623737027004341

Almeida, L. S., Gomes, C. A., Ribeiro, I. S., Dantas, J., Sampaio, M., Alves, M., Rocha, A., Paulo, E., Pereira, T., Nogueira, E., Gomes, F., Marques, L., Sá, C., & Santos, F. (2005). Sucesso e insucesso no ensino básico: relevância de variáveis sócio-familiares e escolares em alunos do 5º ano. In B. D. Silva, L. S. Almeida (Eds.), *Actas do Congresso Galaico-Português de Psicopedagogia* (pp. 3629-2642). Centro de Investigação em Educação do Instituto de Educação e Psicologia da Universidade do Minho.

Anderson, G. E., Jimerson, S. R., & Whipple, A. D. (2005) Student ratings of stressful experiences at home and school: Loss of a parent and grade retention as superlative stressors. *Journal of Applied School Psychology, 21*, 1–20. https://doi.org/10.1300/J370v21n01_01

Barata, M. C., Calheiros, M. M., Patrício, J. N., Graça, J., & Lima, M. L. (2015). Evaluating the impact of national educational policy to reduce retention and increase achievement in compulsory education. *The Elementary School Journal, 116*, 149–171. https://doi.org/10.1086/682748

Bastos, A., & Ferrão, M. E. (2019). Analysis of grade repetition through multilevel models: A study from Portugal. *Cadernos de Pesquisa, 49*, 270–288. https://doi.org/10.1590/198053146131

Belot, M., & Vandenberghe, V. (2014). Evaluating the "threat" effects of grade repetition:

Exploiting the 2001 reform by the French-speaking community of Belgium. *Education Economics, 22*, 73-89. https://doi.org/10.1080/09645292.2011.607266

Bietenbeck, J. (2014). *Learning from adversity? Short- and long-term spillover effects from grade retention in kindergarten*. https://www.researchgate.net/publication/268447733_Learning_from_Adversity_Short-and_Long-Term_Spillover_Effects_from_Grade_Retention_in_Kindergarten

Bonvin, P., Bless, G., & Schuepbach, M. (2008). Grade retention: Decision-making and effects on learning as well as social and emotional development. *School Effectiveness and School Improvement, 19*, 1-19. https://doi.org/10.1080/09243450701856499

Bright, A. D. (2011). *A meta-analysis of the effects of grade retention of K-6 students on student achievement: 1990-2010*. [Doctoral dissertation, University of Alabama]. Institutional Repository of the University of Alabama. http://acumen.lib.ua.edu/content/u0015/0000001/0000747/u0015_0000001_0000747

Brophy, J. (2006). *Grade repetition.* International Academy of Education. International Institute for Educational Planning.

Camacho, A., Correia, N., Zaccoletti, S., & Daniel, J. R. (2021). Anxiety and social support as predictors of student academic motivation during the COVID-19. *Frontiers in Psychology,* 12, 644338. https://doi.org/10.3389/fpsyg.2021.644338

Castella, K., Byrne, D., & Covington, M. (2013). Unmotivated or motivated to fail? A cross-cultural study of achievement motivation, fear of failure, and student disengagement. *Journal of Educational Psychology, 105*, 861–880. https://doi.org/10.1037/a0032464

Conboy, J., Moreira, I., Santos, I., & Fonseca, J. (2013). Práticas e consequências da retenção escolar: Alguns dados do PISA. In L. Veloso & P. Abrantes (Orgs.), *Sucesso escolar: Da compreensão do fenómeno às estratégias para o alcançar* (pp. 9-28). Mundos Sociais.

Conselho Nacional de Educação (2020). *Estado da educação 2019* [State of Education 2019]. Conselho Nacional de Educação. https://www.cnedu.pt/pt/publicacoes/estado-da-educacao/1615-estado-da-educacao-2020

Covington, M. V. (1992). Making the grade: A self-worth perspective on motivation and school reform. Cambridge University Press. https://doi.org/10.1017/CBO9781139173582

Covington, M. V. (2000). Goal theory, motivation and school achievement: An integrative review. *Annual Review of Psychology, 51*, 171-200. https://doi.org/10.1146/annurev.psych.51.1.171

Demanet, J., & Van Houtte, M. (2013). Grade retention and its association with school misconduct in adolescence: A multilevel approach. *School Effectiveness and School Improvement, 24*, 417-434. https://doi.org/10.1080/09243453.2012.727834

Demanet, J., & Van Houtte, M. (2016). Are flunkers social outcasts? A multilevel study of

grade retention effects on same-grade friendships. American Educational Research Journal, 53, 745– 780. https://doi.org/10.3102/0002831216646867

Dupriez, V., Dumay, X., & Vause, A. (2008). How do school systems manage pupils' heterogeneity? *Comparative Education Review, 52*, 245–273. https://doi.org/10.1086/528764

Eccles, J. S., & Midgley, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for early adolescents. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education* (Vol. 3, pp. 139-186). Academic Press.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Iver, D. M. (1993). Development during adolescence. The impact of stage-environment fit on young adolescents' experiences in schools and in families. *The American Psychologist, 48*, 90–101. https://doi.org/10.1037//0003-066x.48.2.90

Ehmke, T., Drechsel, B., & Carstensen, C. H. (2010). Effects of grade retention on achievement and self-concept in science and mathematics. *Studies in Educational Evaluation, 36*, 27–35. https://doi.org/10.1016/j.stueduc.2010.10.003

Ehmke, T., Salzer, C., Pietsch, M., Drechsel, B., & Muller, K. (2017). Competence development in the school year after PISA 2012: Effects of grade retention. *Zeitschrift fur Erziehungswissenschaft, 20*, 99–124. https://doi.org/10.1007/s11618-017-0752-4

Eren, O., Depew, B., & Barnes, S. (2017). Test-based promotion policies, dropping out, and juvenile crime. *Journal of Public Economics, 153,* 9-31. https://doi.org/10.1016/j.jpubeco.2017.07.002

European Commission. (2020*). Equity in school education in Europe. Structures, policies and student performance*. Eurydice report. Publications Office of the European Union. https://doi.org/10.2797/286306

Ferrão, M. E. (2015). Tópicos sobre retenção escolar em Portugal através do PISA: Qualidade e equidade. *Arquivos Analíticos de Políticas Educativas, 23*, 1-19. http://www.redalyc.org/articulo.oa?id=275041389105

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140. http://dx.doi.org/10.1177/001872675400700202

Fine, J. G., & Davis, J. M. (2003). Grade retention and enrollment in post-secondary education. *Journal of School Psychology, 41*, 401-411. http://dx.doi.org/10.1016/j.jsp.2003.07.001

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117–142. https://doi.org/10.3102/00346543059002117

Flores, I., Mendes, R., & Velosa, P. (2013). O que se passa que os alunos não passam. In Conselho Nacional de Educação (Ed.), *Estado da Educação 2013* (pp. 374-391). Conselho Nacional de Educação.

https://www.cnedu.pt/content/edicoes/estado_da_educacao/Estado-da-Educacao-2013-online-v4.pdf

Fraysier, K., Reschly, A., & Appleton, J. (2020). Predicting postsecondary enrolment with secondary student engagement data. *Journal of Psychoeducational Assessment, 38*, 882–899. https://doi.org/10.1177/0734282920903168

Gadeyne, E., Onghena, P., & Ghesquière, P. (2008). child and family characteristics associated with nonpromotion in preprimary education. *Exceptional Children, 74*, 453–469. https://doi.org/10.1177/001440290807400403

García-Pérez, J. I., Hidalgo-Hidalgo, M., & Robles-Zurita, J. A. (2014). Does grade retention affect students' achievement? Some evidence from Spain. *Applied economics, 46*, 1373-1392. http://dx.doi.org/10.1080/00036846.2013.872761

Geng, T., & Rockoff, J. (2017). Does repeating a grade make students (and parents) happier? Regression discontinuity evidence from New York City. https://www.udesa.edu.ar/sites/default/files/rockoff.pdf

Giano, Z., Williams, A. L., & Becnel, J. N. (2022). Grade retention and school dropout: Comparing specific grade levels across childhood and early adolescence. *Journal of Early Adolescence, 42*, 33–57. https://doi.org/10.1177/02724316211010332

Gleason, K. A., Kwok, O. M., & Hughes, J. N. (2007). The short-term effect of grade retention on peer relations and academic performance of at-risk first graders. *The Elementary School Journal, 107*, 327–340. https://doi.org/10.1086/516667

Goos, M., Schreier, B. M., Knipprath, H. M. E., De Fraine, B., Van Damme, J., & Trautwein, U. (2013). How can cross-country differences in the practice of grade retention be explained? A closer look at national educational policy factors. *Comparative Education Review, 57*, 54-84. https://doi.org/10.1086/667655

Goos, M., Van Damme, J., Onghena, P., Petry, K., & de Bilde, J. (2013). First-grade retention in the Flemish educational context: Effects on children's academic growth, psychosocial growth, and school career throughout primary education. *Journal of School Psychology, 51*, 323–347. https://doi.org/10.1016/j.jsp.2013.03.002

Gottfried, M. A. (2013). The spillover effects of grade-retained classmates: Evidence from urban elementary schools. *American Journal of Education, 119*, 405–444. https://doi.org/10.1086/669851

Gubbels, J., van der Put, C. E., & Assink, M. (2019). Risk factors for school absenteeism and dropout: A meta-analytic review. *Journal of Youth and Adolescence 48*, 1637–1667. https://doi.org/10.1007/s10964-019-01072-5

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Higgins, S., Katsipataki, M., Villanueva Aguilera, A. B., Dobson, E., Gascoine, L., Rajab, T.,

Kalambouka, A., Reardon, J., Stafford, J., & Uwimpuhwe, G. (2022). The Teaching and Learning Toolkit: Communicating research evidence to inform decision-making for policy and practice in education. *Review of Education, 10*, e3327. https://doi.org/10.1002/rev3.3327

Holmes, C. T. (1989). Grade-level retention effects: A meta-analysis of research studies. In L. A. Shepard & M. L. Smith (Eds.), *Flunking grades: Research and policies on retention* (pp.16-33). The Falmer Press.

Holmes, C. T., & Matthews, K. M. (1984). The effects of nonpromotion on elementary and junior high school pupils: A meta-analysis. *Review of Educational Research, 54*, 225-236. https://doi.org/10.3102/00346543054002225

Hong, G. L., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis, 27*, 205–224. https://doi.org/10.3102/01623737027003205

Hong, G. L., & Yu, B. (2008). Effects of kindergarten retention on children's social-emotional development: An application of propensity score method to multivariate, multilevel data. *Developmental Psychology, 44,* 407–421. https://doi.org/10.1037/0012-1649.44.2.407

Huang, F. L. (2014). Further understanding factors associated with grade retention: Birthday effects and socioemotional skills. *Journal of Applied Developmental Psychology, 35*, 79-93. https://doi.org/10.1016/j.appdev.2013.12.004

Hwang, S. H., & Cappella, E. (2018). Rethinking early elementary grade retention: Examining long-term academic and psychosocial outcomes. *Journal of Research on Educational Effectiveness, 11*, 559-587. https://doi.org/10.1080/19345747.2018.1496500

Ikeda, M. (2011). PISA in focus: When students repeat grades or are transferred out of School: What does it mean for education systems? http://www.oecd.org/dataoecd/35/58/48363440.pdf

Ikeda, M., & García, E. (2014). Grade repetition: A comparative study of academic and non-academic consequences. *OECD Journal: Economic Studies, 2013*, 1–47. https://doi.org/10.1787/eco_studies-2013-5k3w65mx3hnx

Jackson, G. B. (1975). The research evidence on the effects of grade retention. *Review of Educational Research, 45*, 613-635. https://doi.org/10.2307/1170067

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology, 37*, 243-272. https://doi.org/10.1016/S0022-4405(99)00005-9

Jimerson, S. R. (2001). A synthesis of grade retention research: Looking backward and

moving forward. *California School Psychologist, 6*, 47–59. https://doi.org/10.1007/BF03340883

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). Winning the battle and losing the war: Examining the relation between grade retention and dropping out of high school. *Psychology in the Schools, 39*, 441-457. https://doi.org/10.1002/pits.10046

Jimerson, S., Carlson, E., Rotert, M., Egeland, B., & Sroufe, L. A. (1997). A prospective, longitudinal study of the correlates and consequences of early grade retention. *Journal of School Psychology, 35*, 3-25. https://doi.org/10.1016/S0022-4405(96)00033-7

Jimerson, S. R., & Ferguson, P. (2007). A longitudinal study of grade retention: academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22,* 314-339. https://doi.org/10.1037/1045-3830.22.3.314

Justino, D., Pascueiro, L., Franco, L., Santos, R., Almeida, S., & Batista, S. (2014). *Atlas da Educação: Contextos sociais e locais de sucesso e insucesso: Portugal 1991-2012*. CESNOVA. http://www.epis.pt/downloads/mentores/atlas-da-educacao.pdf

Klapproth, F., & Schaltz, P. (2015). Who is retained in school, and when? Survival analysis of predictors of grade retention in Luxembourgish secondary school. *European Journal of Psychology of Education, 30*, 119–136. https://doi.org/10.1007/s10212- 014-0232-7

Klapproth, F., Schaltz, P., Brunner, M., Keller, U., Fischbach, A., Ugen, S., & Martin, R. (2016). Short-term and medium-term effects of grade retention in secondary school on academic achievement and psychosocial outcome variables. *Learning and Individual Differences, 50*, 182-194. https://doi.org/10.1016/j.lindif.2016.08.014

Kretschmann, J., Vock, M., Lüdtke, O., Jansen, M., & Gronostaj, A. (2019). Effects of grade retention on students' motivation: A longitudinal study over 3 years of secondary school. *Journal of Educational Psychology, 111*, 1432–1446. https://doi.org/10.1037/edu0000353

Lamote, C., Pinxten, M., Van Den Noortgate, W., & Van Damme, J. (2014). Is the cure worse than the disease? A longitudinal study on the effect of grade retention in secondary education on achievement and academic self-concept. *Educational Studies, 40*, 496–514. http://doi .org/10.1080/03055698.2014.936828

Lavy, V., Paserman, M. D., & Schlosser, A. (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *Economic Journal, 122,* 208-237. https://doi.org/10.1111/j.1468-0297.2011.02463.x

Lubbers, M. J., Van der Werf, M. P. C., Snijders, T. A. B., Creemers, B. P. M., & Kuyper, H. (2006). The impact of peer relations on academic progress in junior high. Journal of School Psychology, 44(6), 491-512. https://doi.org/10.1016/j.jsp.2006.07.005

Marsh, H. W. (2016). Cross-cultural generalizability of year in school effects: Negative

effects of acceleration and positive effects of retention on academic self-concept. *Journal of Educational Psychology*, 108, 256–273. https://doi.org/10.1037/edu0000059

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Lichtenfeld, S. (2017). Long-term positive effects of repeating a year in school: Six-year longitudinal study of self-beliefs, anxiety, social relations, school grades, and test scores. *Journal of Educational Psychology, 109*, 425. doi:10.1037/edu0000144

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K. T., O'Mara, A. J., & Craven, R. G. (2008). The Big-fish–little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. *Educational Psychology Review, 20*, 319–350. https://doi.org/10.1007/s10648-008-9075-6

Marsh, H. W., Trautwein, U., Lüdke, O., & Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*, 510-524. https://doi.org/10.1037/0022-0663.100.3.510

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. Journal of Educational Psychology, 89, 41–54. https://doi.org/10.1037/0022-0663.89.1.41

Martin, A. J. (2011). Holding back and holding behind: Grade retention and students' non-academic and academic outcomes. *British Educational Research Journal, 37*, 739–763. https://doi.org/10.1080/01411926.2010.490874

Mathys, C., Véronneau, M.-H., & Lecocq, A. (2019). Grade retention at the transition to secondary school: Using propensity score matching to identify consequences on psychosocial adjustment. *Journal of Early Adolescence, 39,* 97–133. https://doi.org/10.1177/0272431617735651

Mendez, L. M. R., Kim, E. S., Ferron, J., & Woods, B. (2015). Altering school progression through delayed entry or kindergarten retention: Propensity score analysis of long-term outcomes. Journal of Educational Research, 108, 186-203. https://doi.org/10.1080/00220671.2013.867474

Mons, N. (2007*). Les nouvelles politiques éducatives* [New educational policies]. Presses Universitaires de France

Moser, S. E., West, S. G., & Hughes, J. N. (2012). Trajectories of math and reading achievement in low-achieving children in elementary school: Effects of early and later retention in grade. *Journal of Educational Psychology,* 104, 603–621. https://doi.org/10.1037/a0027571

Nascimento, S., & Peixoto, F. (2012). Relações entre o estatuto escolar e o autoconceito,

auto-estima e orientações motivacionais em alunos do 9º ano de escolaridade [Relationship between school status, self-esteem and goal orientations among 9th grade students]. *Análise Psicológica, 30*, 421-434. https://doi.org/10.14417/ap.602

National Association of School Psychologists (2021). COVID-19 resources: Guidance on the use of grade retention and special education eligibility to address instructional loss. https://www.nasponline.org/resources-and-publications/resources-and-podcasts/covid-19-resource-center/return-to-school/guidance-on-the-use-of-grade-retention-and-special-education-eligibility-to-address-instructional-loss

Nunes, L. C., Balcão Reis, A., & Seabra, C. (2018). Is retention beneficial to low-achieving students? Evidence from Portugal. *Applied Economics, 50*, 4306–4317. https://doi.org/10.1080/00036846.2018.1444261

Organisation for Economic Co-operation and Development (2019). *PISA 2018 results: What school life means for students' lives* (Vol. 3). OECD Publishing. https://doi.org/10.1787/acd78851-en

Organisation for Economic Co-operation and Development (2020). *PISA 2018 results: Effective policies, successful schools* (Vol. 5). OECD Publishing. https://doi.org/10.1787/ca768d40-en

Ou, S. R., & Reynolds, A. J. (2010). Grade retention, postsecondary education and public aid receipt. *Educational Evaluation and Policy Analysis, 32*, 118-139. https://doi.org/10.3102/0162373709354334

Özek, U. (2015). Hold back to move forward? Early grade retention and student misbehavior. *Education Finance and Policy, 10*(3), 350-377. https://doi.org/10.1162/EDFP_a_00166

Pagani, L., Tremblay, R. E., Vitaro, F., Boulerice, B., & McDuff, P. (2001). Effects of grade retention on academic performance and behavioral development. *Development and Psychopathology, 13*, 297–315. https://doi.org/10.1017/s0954579401002061

Peixoto, F., & Almeida, L. S. (2010). Self-concept, self-esteem and academic achievement: Strategies for maintaining self-esteem in students experiencing academic failure. *European Journal of Psychology of Education, 25*, 157–175. https://doi.org/10.1007/s10212-010- 0011-z

Peixoto, F., Monteiro, V., Mata, L., Sanches, C., Pipa, J., & Almeida, L. (2016). "To be or not to be retained…That's the question!" Retention, self-esteem, self-concept, achievement goals, and grades. *Frontiers in Psychology, 7*, 1550. https://doi.org/10.3389/fpsyg.2016.01550

Peixoto, F., Pipa, J., Mata, L., Monteiro, V., & Sanches, C. (2017). "To learn, or to be the best?" Achievement goal profiles in pre-adolescents. *Análise Psicológica, 35,* 499-511. https://doi.org/10.14417/ap.1182

Pereira, M. C., & Reis, H. J. (2014). Grade retention during basic education in Portugal:

Determinants and impact on student achievement. *Economic Bulletin, 16,* 61–83. https://www.bportugal.pt/sites/default/files/anexos/papers/ab201407_e.pdf

Range, B. G., Pijanowski, J., Holt, C. R., & Young, S. (2012). The perceptions of primary grade teachers and elementary principals about the effectiveness of grade-level retention. *Professional Educator, 36*. https://eric.ed.gov/?id=EJ989520

Sidali, Z., Pipa, J., & Peixoto, F. (2021, September 1-3). *Determinantes da retenção escolar: Papel das componentes psicossociais* [Poster presentation]. XVI Congresso Internacional Galego-Português de Psicopedagogia, Braga, Portugal.

Santana, M. R. (2019). *Práticas e representações acerca da retenção escolar* [Practices and representations on grade retention] [Doctoral dissertation, Universidade Nova de Lisboa]. Repositório Universidade Nova. https://run.unl.pt/handle/10362/89715.

Santos, N. N., Monteiro, V., & Carvalho, C. (2022). Impact of grade retention and school engagement on student intentions to enrol in higher education in Portugal. *European Journal of Education*. https://doi.org/10.1111/ejed.12535

Santos, N. N., Pipa, J., & Monteiro, V. (2023). Analysing grade retention beliefs within teachers' psycho-pedagogic beliefs system. *Teaching and Teacher Education*, *121.* https://doi.org/10.1016/j.tate.2022.103939

Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics, 152*, 154-169. https://doi.org/10.3386/w21509.

Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: A best-evidence synthesis. Educational Research Review, 6(1), 1-26. https://doi.org/10.1016/j.edurev.2010.07.002

Smith, J., Paquin, S., St-Amand, J., Singh, C., Moreau, D., Bergeron, J., & Leroux, M. (2022). A remediation measure as an alternative to grade retention: A study on achievement motivation. *Psychology in the Schools, 59,* 1209-1221. https://doi.org/10.1002/pits.22677

Smith, M. L., & Shepard, L. A. (1988). Kindergarten readiness and retention: A qualitative study of teachers' beliefs and practices. American Educational Research Journal, 25, 307–333. https://doi.org/10.2307/1163167

Tomchin, E. M., & Impara, J. C. (1992). Unraveling teachers' beliefs about grade retention. *American Educational Research Journal, 29*, 199-223. https://doi.org/10.2307/1162908

Valbuena, J., Mediavilla, M., Choi, A., & Gil, M. (2021). Effects of grade retention policies: A literature review of empirical studies applying causal inference. *Journal of Economic Surveys, 35*, 408-451. https://doi.org/10.1111/joes.12406

Van Canegem, T., Van Houtte, M., & Demanet, J. (2021). Grade retention and academic

self-concept: A multilevel analysis of the effects of schools' retention composition. *British Educational Research Journal, 47,* 1340-1360. https://doi.org/10.1002/berj.3729

Van Canegem, T., Van Houtte, M., & Demanet, J. (2022a). It's a hard-knock life for us : A multilevel analysis on the association between grade retention and being bullied in 25 countries. *Research Papers in Education*. https://doi.org/10.1080/02671522.2022.2125050

Van Canegem, T., Van Houtte, M., & Demanet, J. (2022b). Grade retention: A pathway to solitude? A cross-national multilevel analysis of the effects of being retained on students' sense of belonging. *Comparative Education Review*. https://doi.org/10.1086/721649

Voelkl, K. E. (1996). Measuring students' identification with school. Educational and *Psychological Measurement, 56,* 760-770. https://doi.org/10.1177/0013164496056005003

Wentzel, K. R. (2002). The contribution of social goal setting to children's school adjustment. In A. Wigfield & J. Eccles (Eds.), *Development of achievement motivation* (pp. 222-246). Academic Press.

Wentzel, K. R., Jablansky, S., & Scalise, N. R. (2021). Peer social acceptance and academic achievement: A meta-analytic study*. Journal of Educational Psychology, 113*, 157–180. https://doi.org/10.1037/edu0000468

Wu, W., West, S. G., & Hughes, J. N. (2008). Effect of retention in first grade on children's achievement trajectories over 4 years: A piecewise growth analysis using propensity score matching. *Journal of Educational Psychology, 100*, 727-740. https://doi.org/10.1037/a0013098

Wu, W., West, S. G., & Hughes, J. N. (2010). Effect of grade retention in first grade on psychosocial outcomes. *Journal of Educational Psychology, 102*, 135–152. https://doi.org/10.1037/a0016664

Xia, N., & Kirby, S. N. (2009). *Retaining students in grade: A literature review of the effects of retention on students' academic and nonacademic outcomes*. http://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR678.pdf

Xiang, N., & Chiu, S. (2022). The school matters: Hong Kong secondary schools' grade-retention composition, students' educational performance, and educational inequality. *School Effectiveness and School Improvement.* https://doi.org/10.1080/09243453.2022.2136210

Yamamoto, K., & Byrnes, D. A. (1987). Primary children's ratings of the stressfulness of experiences. *Journal of Research in Childhood Education, 2,* 117–121. https://doi.org/10.1080/02568548709594928

SECTION IV | APPENDICES

APPENDIX A: Supporting information for Chapter 3

Appendix A1

Descriptive Characteristics of the Included Studies by Type of Outcome

Author(s), publication year, type of publication	General Information			Grade retained[a]	Method	Data analysis procedure	Outcome Measures	Results[b]				
	N	Country	Data					G/A	Effect sign			Effect size
									-	ns	+	
Academic achievement												
Alet et al., 2013, journal article	7110 6700	France	1997 nationally representative panel of the French Ministry of Education (followed until 2004)	1-2	IV (birth quarter)	Linear regression models, controlling for several student, class and school characteristics	- Composite score based on standardized national test	G3 G6	0 1	0 0	1 0	-0.32 to 0.52
Appelmans, 2016, master thesis	618-659 E=168-173 C=450-476	Belgium	2003-2014 longitudinal SIBO study in the state of Flanders	1	PS (14 covariates)	Linear regression models, controlling for several student characteristics	- Math IRT calibrated scale score	17y	1	0	0	-0.48
Battistin & Schizzerotto, 2012, report	3176 E=1127 C=2049	Italy	PISA 2006 and 2009 subsample of four provinces in northern region of Italy (Trento, South Tyrol, Verona and Vicenza)	9-10	DID	Linear regression models with standard errors clustered by classes, controlling for several student characteristics	- PISA reading, math and science achievement standardized test score (average of 5 plausible values)	15y	2^	8^	2^	-0.61 to 0.48
Battistin & Schizzerotto, 2019, journal article	1758 E=842 C=916	Italy	Cross-sectional study in four provinces in northern region	10-11	RD	Linear regression models with standard errors	-Reading, math and science achievement	G10-11	2^	10^	0^	-0.43 to 0.20

			of Italy (Trento, South Tyrol, Verona and Vicenza)			clustered by classes, controlling for several student characteristics	standardized test score (PISA derived)					
Belot & Vandenberghe, 2014, journal article	92269 E=7470 C=84799	Belgium (with Great Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden as counter-factuals)	PISA 2003 and 2006 (nationally representative)	7	DID	Linear regression models, controlling for several student and school characteristics, and applying PISA sample weights	- PISA reading, math, science and achievement test score (average of 5 plausible values) and total score (average of previous 3 scores)	15y	0	1	2	-0.04 to 0.09
Bhattacharya, 2008, conference proceeding	4363-4459 E=179 C=4280 E=111 C=4280 E=83 C=4280 E=108 C=4280	US	Combination of 1997 nationally representative NLSY and 1997 NLCS (1986-2002)	1 2 3 4-5	PS (12 covariates)	Linear regression models, controlling for several student characteristics	- PIAT math and reading standardized test score	8y 9y 10y 11-12y	2^ 2^ 1^ 1^	0 0 1^ 1^	0 0 0 0	-0.41 to -0.36 -0.41 to -0.33 -0.23 to -0.13 -0.28 to 0.01
Chen et al., 2010, journal article	1653 E=149 C=1504 E=99 C=1554 E=83 C=1570	China	Retrospective study in 2006 in the rural, extremely poor Shaanxi province	2 3 4	DID and PS (9 covariates)	Linear regression models with township fixed effects, controlling for several student, class, and school characteristics	- Math and language (Chinese) standardized test score	G2-5 G3-5 G4-5	5^,5 0 1^	5^,5 10^,10 9^	0 0 0	-0.27 to 0.01 -0.19 to 0.16 -0.25 to 0.17

Cooley-Fruehwirth et al., 2016, journal article	7668 E=255 C=7038 E=288 C=7038 E=87 C=7038	US	Nationally representative ECLS-K cohort	K 1-2 3-4	FAD	Linear regression models controlling for several student, class and school characteristics	-Math and reading IRT calibrated scale scores	7,9 & 11y 9 & 11y 11y	2^ 0 0	3 2^ 2^	1 2 0	-0.98 to 0.15 -0.15 to 0.19 0.24 to 0.21
d'Haultfoeuille, 2010, journal article	5467	France	1997 nationally representative panel of the French Ministry of Education (2000-2003)	5	IV (previous test score)	Linear regression models	- Composite score based on standardized knowledge evaluation test	G6	1	0	2	-0.43 to 1.17
Diris, 2012, doctoral dissertation	181200	Belgium Canada, Denmark England, Finland, France, Italy, New Zealand, Poland, Portugal, Spain, and Sweden	PISA 2003 and 2009 (nationally representative)	1-6 (SR) 7-8 (SR) 1-8 (SR)	IV (birth month)	Linear regression models with country fixed effects, controlling for some student characteristics	- PISA math, reading and science achievement standardized test score (average of 5 plausible values) and total score (average of previous 3 scores) - Problem solving standardized test score	15y	0 1 1	3 2 0	0 0 0	-0.27 to -0.23 -0.92 to -0.50 -0.63
Diris, 2017, journal article	344551	Austria, Belgium, Canada, Denmark,	PISA 2003, 2009 and 2012 (nationally representative)	1-6 (SR)	IV (birth month)	Linear regression models with country and	- PISA math, reading and science achievement	15y	3	0	0	-0.54 to -0.45

		Estonia, Finland, France, Italy, Luxembourg, the Netherlands, New Zealand, Poland, Portugal, Slovakia, Spain, Sweden				cohort fixed effects, controlling for some student characteristics, and applying PISA sample weights	standardized test score (average of 5 plausible values) and total score (average of previous 3 scores) - Problem solving standardized test score		1	0	0	-0.16
Dong, 2010, journal article	8672 E=281 C=8391	US	nationally representative ECLS-K cohort (1998-2002)	K	IV (age-at-entry and cut-off date)	Linear regression models, controlling for several student, class, and school characteristics	- Math and reading IRT calibrated scale score	G1 & G3	0	1	3	0.18 to 0.55
Ehmke et al., 2010, journal article	720 E=360 C=360	Germany	PISA 2003 and extension (nationally representative, 2002-2004)	9	PS (40 covariates)	Structural equation models with standard errors clustered by schools	- Math and science IRT calibrated scale score	16y	0	2^	0	-0.07 to -0.05
Ehmke et al., 2017, journal article	178 E=89 C=89	Germany	PISA 2012 and extension (nationally representative, 2011-2013)	9	PS (12 covariates)	Structural equation models with standard errors clustered by schools	- Math IRT calibrated scale score	16y	0	1^	0	-0.02

Ferreira Sequeda et al., 2018, report	30576 - 35693	Colombia	2007-2013 multicohort longitudinal study in retention-policy-switching secondary schools (aggregated data at school level)	10	DID	Linear regression models with school and cohort fixed effects, controlling for several school characteristics	- SABER 11 standardized math test score - SABER 11 standardized language (Spanish) test score	G11	0	1	1	0.01 to 0.07
Garcia-Perez et al., 2014, journal article	15662-18687 E=693 C=14969 E=3718 C=14969 E=1036 C=14969	Spain	PISA 2009 (nationally representative)	1-6 7-9 1-9 (twice)	IV (birth quarter)	Switching regression models controlling for several student and characteristics	- PISA math standardized test scores (average of five plausible values)	15y	1 1 1	0 0 0	0 0 0	-0.94 -0.59 -1.06
Gary-Bobo et al., 2016, journal article	12429 E=3026 C=9403	France	1995 nationally representative secondary education Panel and Base Scolarité of the French Ministry of Education	6-8	IV (birth quarter) and FAD	Linear and probit regression models controlling for some student and school characteristics	- Standardized math and language (French) test score - Difference in math and language (French) standardized test scores G9-G6	G9	0	1	1	0.03 to 0.08
Geng & Rockoff, 2017, report	939661 – 939962	US	Four-wave data from the New York City Department of Education	3,5,7+	RD	Linear and probit regression models, controlling for	- Normalized language (English) and math state assessment test score	G3-9	0	0	2^,4	0.36 to 0.79

						some student characteristics						
Glick & Sahn, 2010, journal article	556	Senegal	Combination of nationally representative PASEC and 2003 Senegal Household Education and Welfare survey (1995-2003)	2	IV (previous test score)	Linear regression models with standard errors clustered by schools, controlling for several student, class, and school characteristics	- Math and French tests average score	9y	0	1^	0	-0.2
Goos et al., 2011, conference proceeding	2425 E=120 C=2305	Belgium	2003-2011 longitudinal SiBO study in the state of Flanders	1	PS (58 covariates)	3-level (curvi)linear growth curve models and simple slope analyses	-TR math skills -TR language skills	G1-6	2^,10	2^	0^	-1.03 to -0.03
								8-12y	2^,8	2^	0^	-1.03 to -0.09
Goos et al., 2013, journal article	2425 E=120 C=2305	Belgium	2003-2011 longitudinal SiBO study in the state of Flanders	1	PS (58 covariates)	3-level (curvi)linear growth curve models and simple slope analyses	- (Dutch) three-minutes-test average score	G1-6	7	1^	3^	-1.06 to 0.68
							- Math IRT calibrated scale score	8-12y	4^,6	0	0	-1.74 to -0.84
Greene & Winters, 2007, journal article	71950-73695	US	2000-2004 panels of the Florida Ministry of Education (representative for Florida)	3+	DID	Linear regression models with standard errors clustered by schools, controlling for several student characteristics	- FCAT reading developmental scale score	10-11y	0	0	1^,1	0.05 to 0.40
	7087				fuzzy RD				0	1^	1	0.10 to 0.46

Greene & Winters, 2009, journal article	30109	US	2000-2004 panels of the Florida Ministry of Education (representative for Florida)	3+	fuzzy RD	Linear regression models with standard errors clustered by schools, controlling for several student characteristics	- FCAT reading developmental scale score	11y	0	1	18	-0.03 to 0.70
Hamlin, 2013, doctoral dissertation	68-72 E=34-36 C=34-36	US	1998-2012 panels of the Regent Union School District, an economically privileged area located in Northern Los Angeles County	K-2	PS (u)	Dependent groups t-test	- PRC teacher-rated reading and math achievement - PRC teacher-rated reading and math effort - CST language and math score	G6	0	6	0	-0.22 to 0.06
Hill et al., 2011, journal article	6900 E=233 C=6667	US	Nationally Representative ECLS-K	1	PS (236 covariates)	Bayesian additive regression Trees	- Math and reading achievement standardized test score	9 & 11y	3	1	0	-0.18 to -0.05
Hofer et al., 2010, conference proceeding	86 E=31 C=55	US	Preschool Curriculum Evaluation Research in rural, economically underprivileged areas in Tennessee (2003-2009)	K	PS (27 covariates)	4-level linear regression models, controlling for some student characteristics	- Math, reading, science and social studies test score	G3	0	3	1	0.70 to 1.63
	84 E=24 C=60			2	PS (49 covariates)			G3	0	2	2	0.42 to 0.95

Hong & Raudenbush, 2005, journal article	7631 E=463 C=7168	US	Nationally representative ECLS-K cohort (1998-2000)	K	PS (33 covariates)	2-level linear regression models	- Math and reading IRT calibrated scale score	7y	2^	0	0	-0.67 to -0.66
Hong & Raudenbush, 2006, journal article	654-5385 E=183-231 C=471-5162	US	Nationally representative ECLS-K cohort (1998-2000)	K	PS (213 covariates)	2-level linear regression models	- Math and reading IRT calibrated scale scores	7y	4^	0	0	-0.66 to -0.54
Hong & Yu, 2007, journal article	7631 E=463 C=7168 6267 E=188 C=6079	US	Nationally representative ECLS-K cohort (1998-2004)	K 1	PS (46 covariates) PS (28 covariates)	Multivariate 3-level linear regression models	- Math and reading IRT calibrated q score	7, 9 & 11y 9 & 11y	2^,4 4	0 0	0 0	-0.64 to -0.10 -0.25 to -0.42
Hughes et al., 2010, journal article	321	US	2000-2005 longitudinal study with academically at-risk students in Texas	1	PS (67 covariates)	2-level generalized estimating equation models	- Passage of TAKS reading and math test promotion gate score	G3	0	2	0	0.31 to 0.32
Hwang & Cappella, 2018, journal article	574 E=287 C=287	US	1998-2007 nationally representative ECLS-K cohort	1-2	PS (43 covariates)	Linear regression models controlling for several student and school characteristics	- Math and reading IRT calibrated scale scores - SR and TR of math and reading competence	14y	1	5	0	-0.19 to 0.07
Im et al., 2013, journal article	374 E=75 C=299	US	2000-2010 longitudinal study with academically at-risk students in Texas	1-5	PS (67 covariates)	2-piece 3-level curvilinear growth curve models	- WJ-III math and reading W score	G6-7	0	4	0	-0.05 to 0.06

Jacob & Lefgren, 2004, journal article	7623-8120 4552-5018	US	1997-2000 panels of the Chicago Ministry of Education (representative for Chicago)	3+ 6+	fuzzy RD	Linear regression models with cohort fixed effects, controlling for several student and school characteristics	- ITBS math and reading IRT calibrated scale score	10-11y 13-14y	0 1	1 2^,1	2^,1 0	0.02 to 0.13 -0.08 to 0.03
Keslair, 2007, master thesis	9451 E=1095 C=835	France	1997-2001 nationally representative panel of the French Ministry of Education	1-2	IV (rank in class) and RD	Linear regression models, controlling for some student characteristics	- Math standardized test score	G3	0	12	0	-1.21 to 1.02
Klapproth et al., 2016, journal article	136 E=68 C=68 278 E=139 C=139	Luxembourg	2009-2013 longitudinal study (nationally representative)	7 8	PS (16 covariates)	ANOVA models	- Math, German and French school mark - Math, German and French standardized test score	G7-9 G8-9	0 0	7 6	3^,2 3^	-0.10 to 1.11 -0.24 to 1.35
Koppensteiner, 2014, journal article	244081	Brazil	Combination of 2-cohort 2000-2006 longitudinal Programa de Avaliação da Educação Básica in the economically privileged state of Minas Gerais and school census data provided by the	1-4	DID	Linear and logistic regression models with school fixed effects, controlling for several student, class and school characteristics	- Math standardized test score	G4	0	0	1	0.06

			Federal Ministry of Education									
Lamote et al., 2014, journal article	1596 E=215 C=1380	Belgium	1990-1997 longitudinal LOSO study in state of Flanders	8	PS (32 covariates)	3-level curvilinear growth curve models	- (Dutch) language achievement IRT calibrated test score	G8,10 & 12	1	1^,1	0	-0.92 to -0.01
Lorence, 2014, journal article	38731 E=1249 C=37482	US	1993-2002 longitudinal study with academically at-risk students in Texas public schools (representative for Texas)	3+	PS (41 covariates)	2-level linear regression models	- TAAS reading test score	G3-10	0	0	1^,6	0.38 to 1.55
Lorence & Dworkin, 2006, journal article	4025-15684 E=129-449 C=3896-15235	US	1993-2002 longitudinal study with academically at-risk students in Texas public schools (representative for Texas)	3+	PS (21 covariates)	ANCOVA models, controlling for several student, school, and district characteristics	- TAAS reading test score	G3-10	0	3	3^,15	0.08 to 1.96
Mahjoub, 2017, journal article	13136 E=2339 C=10797	France	1995-2001 longitudinal DEPP study with a nationally representative sample (Panel 95)	6-8	IV (birth quarter)	Linear regression models, controlling for several student characteristics	- National standardized math and language (French) test G9-G6 difference score	G9	0	0	2	1.28 to 1.43
Mariano & Martorell, 2013, journal article	394-3315	US	2004-2008 2-cohort longitudinal	5+	fuzzy RD	Polynomial regression models, controlling for	- State assessment test language (English) and	G6-7	0	2	22	0.17 to 0.58

			study in New York City			some student characteristics	math scale score					
McCombs et al., 2009, report	2043-2941 E=589-1293 C=1322-2211	US	2003-2008 4-cohort longitudinal study in New York City	3+	fuzzy RD	Polynomial regression models	- State assessment language (English) and math test scale score	G4-5	0	0	6	0.48 to 0.81
Mendez et al., 2015, journal article	6707 E= 403 C=6273	US	1989-2002 longitudinal Omnibus Project in one large, urban school district in the state of Florida	K	PS (12 covariates)	Logistic and linear regression models, with data split by student lunch status	- CTBS-4 reading, language and math standardized score	G3 & G5 & G7	18	0	0	-1.07 to -0.53
Moser et al., 2012, journal article	363 E=112 C=251	US	2000-2007 longitudinal study with academically at-risk students in Texas	1	PS (72 covariates)	3-level (curvi)linear growth curve models and simple slope analyses	- WJ-III math and reading W score	G1, G2 & G5	0	2	2^,2	-0.10 to 0.97
Nunes et al., 2018, journal article	1240-2825	Portugal	2006-2010 longitudinal study among nationally representative sample	4	PS (9 covariates) and IV (percentage of repeaters in school)	Linear regression models, controlling for several student characteristics	- Math and language (Portuguese) standardized national exam score	G6	0	3	5	0.02 to 0.26
Park, et al., 2018, journal article	4429 E=286 C=4143	US	1998-2002 ECLS-K cohort (nationally representative)	K	PS (42 covariates)	Linear regression model	- Math IRT calibrated scale score	9y	1	0	0	-0.21

Pereira & Reis, 2014, journal article	10906	Portugal	PISA 2003 and 2009 (nationally representative)	1-6 (SR)	IV (maturity and average level of regional retention)	Linear regression models with country and cohort fixed effects, controlling for some student, school and region characteristics, and applying PISA sample weights	- PISA reading and math Standardized test scores (average of five plausible values):	15y	2	0	0	-1.33 to -1.15
				7-9 (SR)					0	0	2	0.21 to 0.24
	279936	Austria, Belgium, Bulgaria, the Czech Republic, Denmark, Estonia, Germany, Finland, France, Greece, Hungary,		1-6 (SR)				15y	2	0	0	-0.90 to -0.75

		Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden	7-9 (SR)					0	0	2	0.49 to 0.57
Roderick & Nagaoka, 2005, journal article	11625	US	1998-2000 panels from the Chicago Public Schools (representative for Chicago)	3+	DID	3-level growth curve models, controlling for several student characteristics	- ITBS calibrated scale scores in reading	10y	0	0	1^	0.02
	6214				RD			10-11y	0	3	1^	0.01 to 0.03
	3257			6+				13-14y	1^,3	0	0	-0.27 to -0.04
Schumann, 2012, report	136	Senegal	1995-2000 longitudinal PASEC (representative for Senegal)	2	PS (10 covariates)	ANOVA	- Math Warm likelihood estimates test score	9-12y	1^,1	2	0	-0.17 to -0.07
	138			3				10-12y	1^	2	0	-0.28 to -0.02
	52			4				11-12y	1^,1	0	0	-0.35 to -0.32
	143			5				12y	1^	0	0	-0.22
Schwerdt, et al., 2017, journal article	22977-76208	US	2000-2005 longitudinal study with	3+	fuzzy RD	Linear and probit regression	- FCAT reading and math	G4-G10	0	0	12	0.13 to 0.73

			academically at-risk students in Texas			models with *SE*s clustered by schools, controlling for several student characteristics	developmental scale score	10-16y	2	3	8	-0.08 to 0.49
Troncin, 2005, doctoral dissertation	206 E=103 C=103	France	2002-2004 longitudinal study in the Côte-d'Or department	1	PS (13 covariates)	4-level linear regression models	- Math, French, science, processing test score and total score (average of previous 4 scores)	8y	4^	0	0	-0.42 to -0.23
Uysal, 2010, conference proceeding	1643 E=303 C=1340	Germany	1970-1980 longitudinal study in North Rhine-Westphalia, an economically privileged area in West Germany	10-12	PS (28 covariates)	Doubly robust linear regression models	- Average grade in math and language (German)	G12	3	0	0	-0.41 to -0.25
Vandecandelaere et al., 2011, report	792-1704 E=315 C= 1389 E=114 C=678	Belgium	2002-2009 longitudinal SiBO study in the state of Flanders	K	PS (39 covariates)	3-level (curvi)linear growth curve models and simple slope analyses	- Math IRT calibrated scale score	K-G1 K-G6	0 0	1 5	1^ 1^,1	0.02 to 0.43 0.00 to 1.23
Vandecandelaere et al., 2013, journal article	1704 E=315 C=1389	Belgium	2002-2008 longitudinal SiBO study in the state of Flanders	K	PS (39 covariates)	3-level (curvi)linear growth curve models	- Math IRT calibrated scale score	7-10y	1^,1	0	0	-1.17 to -0.52

Vandecandelaere et al., 2016b, journal article	524 E=257 C=267	Belgium	2002-2009 longitudinal SIBO study in the state of Flanders	K	PS (28 covariates)	Generalized estimating equation models	- Math IRT calibrated scale scores	7-12y	1^,4	1	0	-1.32 to -0.37
	502 E=235 C=267			1				8-12y	1^,4	0	0	-0.84 to -0.66
Vandecandelaere et al., 2016c, journal article	458 E=61 C=397	Belgium	2002-2009 longitudinal SiBO study in the state of Flanders	2	PS (22 covariates)	Generalized estimating equation model, with time clustered within students and contrast analyses	- Math IRT calibrated scale score	9-12y	3	1^	0	-0.69 to -0.25
Winters & Greene, 2012, journal article	30808	US	2002-2008 panels of the Chicago Ministry of Education (representative for Chicago)	3+	fuzzy RD	Linear regression models with school fixed effects and standard errors clustered by schools, controlling for several student characteristics	- FCAT and Stanford 10 reading, math and science standardized scale score	G4-8	0	3	35	0.01 to 0.52
Wu et al., 2008, journal article	196 E=98 C=98	US	2001-2005 longitudinal study with academically at-risk students in Texas	1	PS (72 covariates)	2-piece 3-level linear growth curve models	- WJ-III math and reading grade standard score	G1-3	0	0	2^,4	0.31 to 0.75
							- WJ-III math and reading W score	8-10y	2^,4	0	0	-0.47 to -0.07
Psychosocial functioning												

Cham et al., 2015, journal article	561 E=177 C=384	US	2001-2011 longitudinal study with academically at-risk students (score school district language test < median) in Texas	1-5	PS (67 covariates)	Multi-group structural equation model	-SR general motivation for educational attainment factor scores (teacher and peer educational aspirations; value of education)	G9	0	1	3	-0.25 to 1.00
Detienne, 2017, master thesis	2743-2829 E=61-62 C=2682-2767	Belgium	2013-2016 longitudinal SiBO study in the state of Flanders	7	PS (18 covariates)	Linear regression models, controlling for several student characteristics	- SR school well-being	14-15y	0	0	1^,1	0.26 to 0.38
Diris, 2012, doctoral dissertation	181200	Belgium Canada, Denmark England, Finland, France, Italy, New Zealand, Poland, Portugal, Spain, and Sweden	PISA 2003 and 2009 (nationally representative)	1-8 (SR)	IV (birth month)	Linear regression models with country fixed effects, controlling for some student characteristics	- SR attitudes towards school, teacher, school subjects, peers, lessons, education confidence towards math topics, math skills; overall confidence and attitudes)	15y	0	9	0	-0.52 to 0.50

Diris, 2017, journal article	344551	Austria, Belgium, Canada, Denmark, Estonia, Finland, France, Italy, Luxembourg, the Netherlands, New Zealand, Poland, Portugal, Slovakia, Spain, Sweden	PISA 2003, 2009 and 2012 (nationally representative)	1-6 (SR)	IV (birth month)	Linear regression models with country and cohort fixed effects, controlling for some student characteristics, and applying PISA sample weights	-SR attitudes towards school, teacher, math; sense of belonging; math self-efficacy; math anxiety)	15y	0	7	0	-0.15 to 0.27
Ehmke et al., 2010, journal article	720 E=360 C=360	Germany	PISA 2003 and extension (nationally representative, 2002-2004)	9	PS (40 covariates)	Structural equation models with standard errors clustered by schools	-SR math self-concept	16y	0	0	1^	0.30
Ehmke et al., 2013, conference proceeding	720 E=360 C=360	Germany	PISA 2003 and extension (nationally representative, 2002-2004)	9	PS (40 covariates)	Structural equation models with standard errors clustered by schools	- SR math Anxiety, boredom, joy; satisfaction with math teacher; self-efficacy in math; self-concept in Math; aggressive orientation in peer group; member of a peer group)	16y	1^	3^	3^	-0.28 to 0.26

Ehmke et al., 2017, journal article	178 E=89 C=89	Germany	PISA 2012 and extension (nationally representative, 2011-2013)	9	PS (12 covariates)	Structural equation models with standard errors clustered by schools	-SR interest in math; instrumental motivation in math; work ethics in math; behaviour in math; self-efficacy in math	16y	1^	4^	0	-0.18 to 0.34
Eren et al., 2017., journal article	23845-25406	US	1999-2012 panels of the Louisiana Department of Education and Department of Public Safety and Corrections, Youth Services, Office of Juvenile Justice (representative for Louisiana)	4	fuzzy RD and IV (July test cut-off)	Linear regression models with standard errors clustered by schools, controlling for some student characteristics	- Juvenile crime (and classification)	17y	0	4	0	-0.01 to 0.00
	37272-39007			8			- Juvenile crime (and classification)	17y	0	4	0	0.00
Geng & Rockoff, 2017, report	163594 – 307565	US	Four-wave data from the New York City Department of Education	3,5,7+	RD	Linear and probit regression models, controlling for some student characteristics	- PR and SR satisfaction (overall; school safety; school environment)	G6-9	0	1^,6	1^,1	0.00 to 0.01
Goos et al., 2013, journal article	2344 E=121 C=2223	Belgium	2003-2011 longitudinal SiBO study in the state of Flanders	1	PS (52 covariates)	3-level (curvi)linear growth curve models and simple slope analyses	- TR peer relationships; aggressive behaviours; hyperactivity; asocial behaviours; independent behaviour in	G1-6	4^,19	10^,16	0	-2.58 to 087

							class; school wellbeing; self-confidence)	8-12y	3^,15	10^,13	1^	-2.35 to 0.44
Hofer et al., 2010, conference proceeding	86 E=31 C=55	US	Preschool Curriculum Evaluation Research in rural, economically underprivileged areas in Tennessee (2003-2009)	K	PS (27 covariates)	4-level linear regression models, controlling for some student characteristics	-TR behaviour (aggressive; prosocial; withdrawn; self-regulation)	G3	0	4	0	-0.07 to 0.36
	84 E=24 C=60			2			-TR behaviour (aggressive; prosocial; withdrawn; self-regulation)	G3	0	3	1	0.20 to 0.70
Hong & Yu, 2008, journal article	7631 E= 463 C=7148	US	Nationally representative ECLS-K cohort (1998-2004)	K	PS (46 covariates)	Multivariate 3-level linear regression models	-SR interest and competence in reading, math and all subjects; competence in peer relationships; internalized problems - TR and PR internalized problems	7, 9 & 11y	0	1^,9	1^,3	-0.14 to 0.17
Hwang & Cappella, 2018, journal article	574 E=287 C=287	US	1998-2007 nationally representative ECLS-K cohort	1-2	PS (43 covariates)	Linear regression models controlling for several student and school characteristics	- SR social self-concept; internalizing behaviours; locus of control; self-esteem	14y	0	3	1	-0.02 to 0.20

Im et al., 2013, journal article	374 E=75 C=299	US	2000-2010 longitudinal study with academically at-risk students in Texas	1-5	PS (67 covariates)	2-piece 3-level curvilinear growth curve models	- TR behavioural and affective engagement	G6-8	0	4	0	-0.09 to 0.11
Klapproth et al., 2016, journal article	136 E=68 C=68	Luxembourg	2009-2013 longitudinal study (nationally representative)	7	PS (16 covariates)	ANOVA models	- SR academic self-concept; school-related anxiety; interest in school; satisfaction with school; teacher-student relationship; instruction disruption; class climate	G7-9	1	6	0	-0.34 to 0.06
	278 E=139 C=139			8				G8-9	0	7	0	-0.22 to 0.14
Kretschmann et al., 2019, journal article	1139 E=61 C=1078	Germany	(2007-2011) longitudinal BiKS-8–14 Study in the states of Bavaria and Hesse	6	PS (18 covariates)	Linear regression models, controlling for several student characteristics	-SR German self-concept, math self-concept, interest in German, interest in math)	13-14y	2^	2^,2	0	-0.46 to 0.14
								G6-7	1^	3^,2	0	-0.49 to 0.15
Lamote et al., 2014, journal article	1596 E=215 C=1380	Belgium	1990-1997 longitudinal LOSO study in state of Flanders	8	PS (32 covariates)	3-level curvilinear growth curve models	- SR academic self-concept	G8,10 & 12	0	2	1^	-0.23 to 0.81
Mathys et al., 2019, journal article	181 E=36 C=145	Belgium	2012 2-wave longitudinal study in one secondary school in the	7	PS (26 covariates)	Univariate ANCOVAs	- SR self-esteem; intrinsic motivation (knowledge, accomplish-	13y	11^	6^	0	-0.80 to 0.20

			state of Wallonia				ment, stimulation); extrinsic motivation (identified regulation, introjected regulation, external regulation); amotivation; Behaviour problems (social withdrawal; anxiety and depression; attention problems, aggressive behaviours; delinquent behaviours; parental support (support for autonomy, support for competence, involvement in the relationship); Class climate victim of violence					
Mendez et al., 2015, journal article	6707 E= 403 C=6273	US	1989-2002 longitudinal Omnibus Project in one	K	PS (12 covariates)	Logistic and linear regression models, with	- TR behaviour (overall behaviour;	G3 & G5	8	8	0	-1.84 to 0.21

			large, urban school district in the state of Florida			data split by student lunch status	difficulty in paying attention; attitudes towards school); - SR adjustment (connection to teachers, connection to school)					
Özek, 2015, journal article	21712-178248	US	2003-2009 panels of the Florida Ministry of Education	3+	fuzzy RD	Polynomial and linear regression models with cohort fixed effects, controlling for several student characteristics and one school characteristic	- Number of disciplinary incidents	10-14y	2^,2	6	0	-0.28 to 0.05
	21712-87924						- Number of disciplinary incidents	G3-8	5	5	0	-0.24 to -0.03
Vandecandelaere et al., 2016a, journal article	1536 E=99 C=1437	Belgium	2002-2009 longitudinal SIBO study in the state of Flanders	K	PS (24 covariates)	3-level multilevel regression models	- TR behaviour (independent, cooperative, hyperactivity, aggression, asocial) - TR attitudes to work; self-confidence -TR peer relationships - TR wellbeing	7-12y	0	3^,15	5^,18	-0.37 to 0.67

Wu et al., 2010, journal article	375 E=124 C=251	USA	2001-2005 longitudinal study with academically at-risk students in Texas	1	PS (72 covariates)	2-piece 3-level linear growth curve models	- TR and PR emotional problems; behavioural engagement; social acceptance	8-10y	0	4^,9	4^7	-0.32 to 0.65
	E=124 C=251					2-level linear regression models	- SR school belonging; academic self-efficacy)		0	1^,3	1^,1	0.02 to 0.59
School career												
Afsa, 2011, report	14987	France	1995 nationally representative panel of the French Ministry of Education (followed until 2003)	6	IV (rank in class)	Probit regression models, controlling for several student characteristics and some school characteristics	- Attainment of secondary education diploma (regardless of track)	G10 or G12	0	1	0	-0.20
Allensworth, 2005, journal article	113937	US	1992-1998 panels of the Chicago ministry of education (representative for Chicago)	8+	DID	Two-level hierarchical generalized linear models, controlling for several student characteristics	- High-school dropout	17y	1	0	0	-0.58
André, 2009, report	1818-1823	Senegal	Combination of nationally representative PASEC and 2003 Senegal Household Education and Welfare survey (1995-2003)	2-6	IV (teacher attitudes toward retention) and RD	Probit regression models with grade and cohort fixed effects, controlling for several student and class characteristics	- Primary school dropout	9-13y	2	0	0	-0.89 to -0.50

Andrew, 2014, journal article	4352 E=424 C=3928	US	1986-2010 NLSY 1979 (Child and Young Adult supplements)	K-5 (SR)	PS (16 covariates)	Logistic regression models, controlling for several student characteristics	- High-school graduation - Entry into postsecondary education	21-40y	2	0	0	-0.56 to -0.35
	8808 E=848 C=7960		1988-2000 NELS 1988		PS (10 covariates)		- High-school graduation - Entry into postsecondary education - Completion of a bachelor's degree	26y	3	0	0	-0.77 to -0.47
Belot & Vandenberghe, 2014, journal article	92269 E=7470 C=84799	Belgium (with Great Britain, Greece, Hungary, Ireland, Iceland, Norway, Poland and Sweden as counter-factuals)	PISA 2003 and 2006 (nationally representative)	7	DID	Linear regression models, controlling for several student and school characteristics, and applying PISA sample weights	- Grade 10 attainment	15y	1	0	0	-0.02
Cockx et al., 2019, report	3993 E=874 C=3059	Belgium	Retrospective study in 2001 and 2003 in the state of Flanders (regionally representative)	8-10	FAD	Discrete choice models	- Evaluation (A, B, C)	G8-10	0	1^,1	0	0.00
							-Track downgrading (general, technical, vocational	G9 & G12	4	2	0	-0.10 to -0.01
							- Delay entry last school year	G12	3	0	0	-0.50 to -0.48

							- High-school diploma	23y	3	0	0	-0.08 to -0.07
Eide & Showalter, 2001, journal article	477-3291	US	1980-1992 longitudinal US High School and Beyond survey	K-10 (SR)	IV (K entry date)	Linear regression models with standard errors clustered by schools, controlling for some student characteristics	-High-school dropout	18-20y	0	4	0	-0.21 to 0.45
Eren et al., 2017., journal article	23845-25406	US	1999-2012 panels of the Louisiana Department of Education and Department of Public Safety and Corrections, Youth Services, Office of Juvenile Justice (representative for Louisiana)	4	fuzzy RD and IV (July test cut-off)	Linear regression models with standard errors clustered by schools, controlling for some student characteristics	- High-school dropout	18y	0	2	0	-0.03 to 0.00
	37272-39007			8			- High-school dropout	18y	2	0	0	
Gary-Bobo et al., 2016, journal article	12429 E=3026 C=9403	France	1995 nationally representative secondary education Panel and Base Scolarité of the French Ministry of Education	6-8	IV (birth quarter) and FAD	Linear and probit regression models controlling for some student and school characteristics	- Probability of access to grade 9	G9	1	0	0	-0.06
Geng & Rockoff, 2017, report	945555 – 1021380	US	Four-wave data from the New York City Department of Education	3,5,7+	RD	Linear and probit regression models, controlling for	- Absences; suspension; placement in special education;	G3-8	1^,4	1	0	-0.53 to -0.03

						some student characteristics	grade retention					
Glick & Sahn, 2010, journal article	664	Senegal	Combination of nationally representative PASEC and 2003 Senegal Household Education and Welfare survey (1995-2003)	2	IV (previous test score)	Ordered probit regression model, controlling for several student, class, and school characteristics	- Primary school dropout	G4	1	0	0	-0.27
Goos et al., 2013, journal article	3558 E=285 C=3273	Belgium	2003-2011 longitudinal SiBO study in the state of Flanders	1	PS (74 covariates)	2-level logistic regression models	- Track choice (high versus low) - Grade retention; placement in special education; school mobility	G7 8-12y	1 2	0 0	0 1	-0.82 -0.53 to 1.20
Hughes et al., 2017a, journal article	538 E=171 C=367	US	2000-2013 longitudinal study with academically at-risk students in Texas	1-5	PS (65 covariates)	Logistic regression model, with standard errors clustered by schools, controlling for some student characteristics	- High-school dropout	16-18y	1	0	0	-0.29
Hughes et al., 2017b, journal article	734 E=256 C=478	US	2000-2016 longitudinal study with academically at-risk students in Texas	1-5	PS (65 covariates)	Multinomial logistic regression model	- High-school completion (high-school diploma, GED, dropout; with high school	20y	1	1	0	-0.54 to -0.53

							diploma as reference category)					
Jacob & Lefgren, 2009, journal article	7754-11777	US	1997-2000 panels of the Chicago Ministry of Education (representative for Chicago)	6+	fuzzy RD	Logistic regression models with standard errors clustered by cohorts, controlling for several student characteristics	- Grade retention - Age at start senior high school - High-school dropout - School mobility - Age at high-school dropout - Number of credits obtained upon high-school dropout	G8 G9 18-22y	0 1 0	0 0 13	1 0 0	0.05 -0.06 -0.09 to 0.02
	2424-3599			8+			- Age at start senior high school - High-school dropout - School mobility - Age at high-school dropout - Number of credits obtained upon high-school dropout	G9 18-22y	2 10	0 16	0 0	-0.79 to -0.40 -0.15 to 0.02
Koppensteiner, 2014, journal article	244081	Brazil	Combination of 2-cohort 2000-2006 longitudinal	1-4	DID	Linear and logistic regression models with	- Primary school dropout - School mobility	G1-G4	1	11	0	0.00

			Programa de Avaliação da Educação Básica in the economically privileged state of Minas Gerais and school census data provided by the Federal Ministry of Education			school fixed effects, controlling for several student, class and school characteristics						
Leighton et al., 2019, journal article	u	Brazil	1999-2006 Panels of the Censo Escolar from the Instituto Nacional de Estudos e Pesquisas Educacionais	1-4	DID	Linear regression models with school fixed effects, controlling for some school characteristics	- School delay	9-12y	4	0	0	-0.02 to -0.01
							- School enrolment	8-12y	0	5	0	-0.01 to 0.00
Lorence, 2014, journal article	38731 E=1249 C=37482	US	1993-2002 longitudinal study with academically at-risk students in Texas public schools (representative for Texas)	3+	PS (41 covariates)	2-level logistic regression models	- High-school enrolment - High-school test taking	G10	0	1	1	0.00 to 0.08
Mahjoub, 2017, journal article	13136 E=2339 C=10797	France	1995-2001 longitudinal DEPP study with a nationally representative sample (Panel 95)	6-8	IV (birth quarter)	Linear regression models, controlling for several student characteristics	- Junior high-school graduation	G10	0	0	1	0.01

Manacorda, 2012, journal article	76438	Uruguay	1996-2001 4-cohort longitudinal study in public nonvocational school	7-9	fuzzy RD	Polynomial regression models with standard errors clustered by schools, grades, and cohorts, controlling for some student characteristics	- Number of grades attended - High-school dropout (from public non-vocational education)	G7-12	6	3	0	-0.27 to 0.15
Martorell & Mariano, 2018, journal article	4753-25898	US	2004-2011 multi-cohort longitudinal study in New York City	3+	fuzzy RD	Polynomial regression models, controlling for some student characteristics	- Suspensions (0/1, number of days suspended); attendance rate; chronic absences	10-12y	0	10^,20	2^,4	-0.01 to 0.04
				4+				11-13y	0	12^,23	1	-0.03 to 0.05
				5+				12-14y	0	6^,24	6^	-0.02 to 0.04
				6+				13-15y	2	12^,22	0	-0.05 to .0.04
				7+				14-16y	3	11^,16	1^,5	-0.07 to 0.07
				8+				15-17y	4^	8^,24	0	-0.08 to 0.02
Mendez et al., 2015, journal article	6707 E= 403 C=6273	US	1989-2002 longitudinal Omnibus Project in one large, urban school district in the state of Florida	K	PS (12 covariates)	Logistic and linear regression models, with data split by student lunch status	- Grade retention; placement in special education; office referrals; high-school graduation timing	G1-12	14	6	0	-1.75 to 0.19
Moser et al., 2012, journal article	363 E=112 C=251	US	2000-2007 longitudinal study with academically at-	1	PS (72 covariates)	logistic regression models	- Grade retention; placement in	G1-5	0	1	1	0.02 to 0.26

			risk students in Texas				special education)					
Okurut, 2018, journal article	30053 26720	Uganda	combination of 2004 and 2010 National Assessment of Progress in Education and 2004 and 2010 Uganda National Household Survey (nationally representative)	1-3 1-5	PS and DID	Probit regression models, controlling for several school, region, teacher and student characteristics	- School dropout	G3 G6	0 0	0 1	1 0	0.04 0.00
Ou & Reynolds, 2010, journal article	917-1216 E=176-210 C=741-1006 814-1073 E=80-86 C=734-987	US	1986-1995 longitudinal study among (a representative sample of) low-income minority children in Chicago	1-3 4-8	PS (15 covariates)	Probit regression models, controlling for some student characteristics	- Tertiary education attendance	24y	0 3	3 0	0 0	-0.04 to -0.03 -0.15 to 0.09
Özek, 2015, journal article	21712-178248 21712-87924	US	2003-2009 panels of the Florida Ministry of Education	3+	fuzzy RD	Polynomial and linear regression models with cohort fixed effects, controlling for several student characteristics and one school characteristic	- Number of in-school suspensions - Number of out-of-school suspensions	10-14y G3-8	2^,2 7	2^,13 13	1 0	-0.12 to 0.10 -0.13 to 0.04

Peterson & Hughes, 2011, journal article	769 E=165 C=604	US	2001-2003 longitudinal study with academically at-risk students in Texas	1	PS (72 covariates)	Structural equation models with standard errors clustered by schools	- Number of educational services received (reduced class size; one-to-one adult tutoring; peer tutoring; remedial instruction outside of the classroom; small-group intensive tutoring, remedial instruction before or after school; one-to-one tutoring by an adult before or after school)	8y	1^	0	0	-0.02
Schwerdt, et al., 2017, journal article	9816-7644	US	2000-2005 longitudinal study with academically at-risk students in Texas	3+	fuzzy RD	Linear and probit regression models with *SE*s clustered by schools, controlling for several student characteristics	- Grade retention; grade progression; high-school entry; high-school graduation; high-school completion of highest grade; high-school leaving age; high-school	11-20y	8	19	8	-0.50 to 0.89

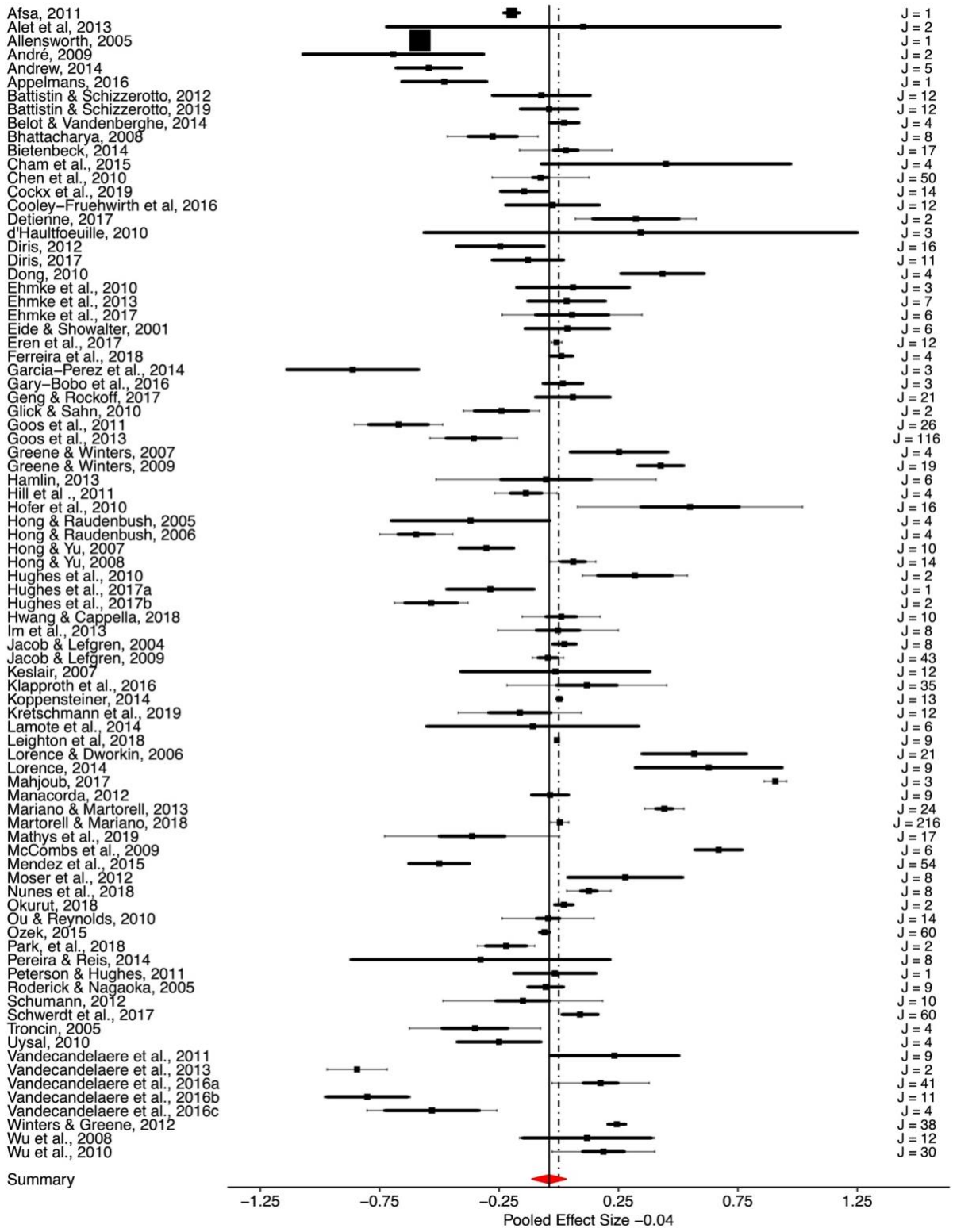
							GPA; high-school course-taking (remedial courses, college prep); high-school total credits; college enrolment)					
Uysal, 2010, conference proceeding	2726 E=520 C=2226	Germany	1970-1980 longitudinal study in North Rhine-Westphalia, an economically privileged area in West Germany	10-12	PS (28 covariates)	Doubly robust generalized linear regression models	- Attainment of Abitur (i.e., diploma academic track)	G12	1	0	0	-0.01
Job career												
Eide & Showalter, 2001, journal article	7809	US	1980-1992 longitudinal US High School and Beyond survey	K-10 (SR)	IV (K entry date)	Linear regression models with standard errors clustered by schools, controlling for some student characteristics	-Log of average year income	28-30y	0	2	0	0.01 to 0.06
Ou & Reynolds, 2010, journal article	1165 E=211 C=954	US	1986-1995 longitudinal study among (a	1-3	PS (15 covariates)	Probit regression models,	- Public aid receipt	24y	0	4	0	-0.05 to -0.03

	1021 E=84 C=937		representative sample of) low-income minority children in Chicago	4-8		controlling for some student characteristics			0	4	0	-0.04 to 0.00
Peers' development												
Bietenbeck, 2014, report	4144-5614 E=102-175 C=4144-5617	US	Longitudinal Student-Teacher Achievement Ratio study in low-income neighbourhoods in Tennessee (1985-1998)	K	RCT	Linear regression models with school fixed effects, controlling for some student characteristics and one class characteristic	- SAT and CTBS math, reading standardized test score of non-retained peers	K, G1, G5 & G8	1^	1^,6	0	-0.09 to 0.06
	1628-1731 E=32-37 C=1628-1731						-TR Behaviour standardized test score (effort; initiative; discipline) of non-retained peers	G4 & G8	0	3	3	0.02 to 0.19
	2438-6039 E=40-193 C=2438-6039						-High-school GPA, graduation; college entrance exams (ACT or SAT) of non-retained peers	G12	0	2	1	0.02 to 0.36
Ferreira Sequeda et al., 2018, report	30576-35693	Colombia	2007-2013 multicohort longitudinal study in retention-policy-switching secondary	10	DID	Linear regression models with school and cohort fixed effects, controlling for	- SABER 11 standardized math test score of non-retained peers - SABER 11 standardized		0	2	0	-0.05 to 0.01

			schools (aggregated data at school level)			several school characteristics	language (Spanish) test score of non-retained peers					
Hong & Raudenbush, 2005, journal article	7631 E=463 C=7168	US	Nationally representative ECLS-K cohort (1998-2000)	K	PS (33 covariates)	2-level linear regression models	- Math and reading IRT calibrated scale score of non-retained peers	7y	0	2^	0	-0.09 to -0.06
Park, et al., 2018, journal article	4429 E=286 C=4143	US	1998-2002 ECLS-K cohort (nationally representative)	K	PS (42 covariates)	Linear regression model	- Math IRT calibrated scale score of non-retained peers		0	0	1	-0.23

Appendix A2

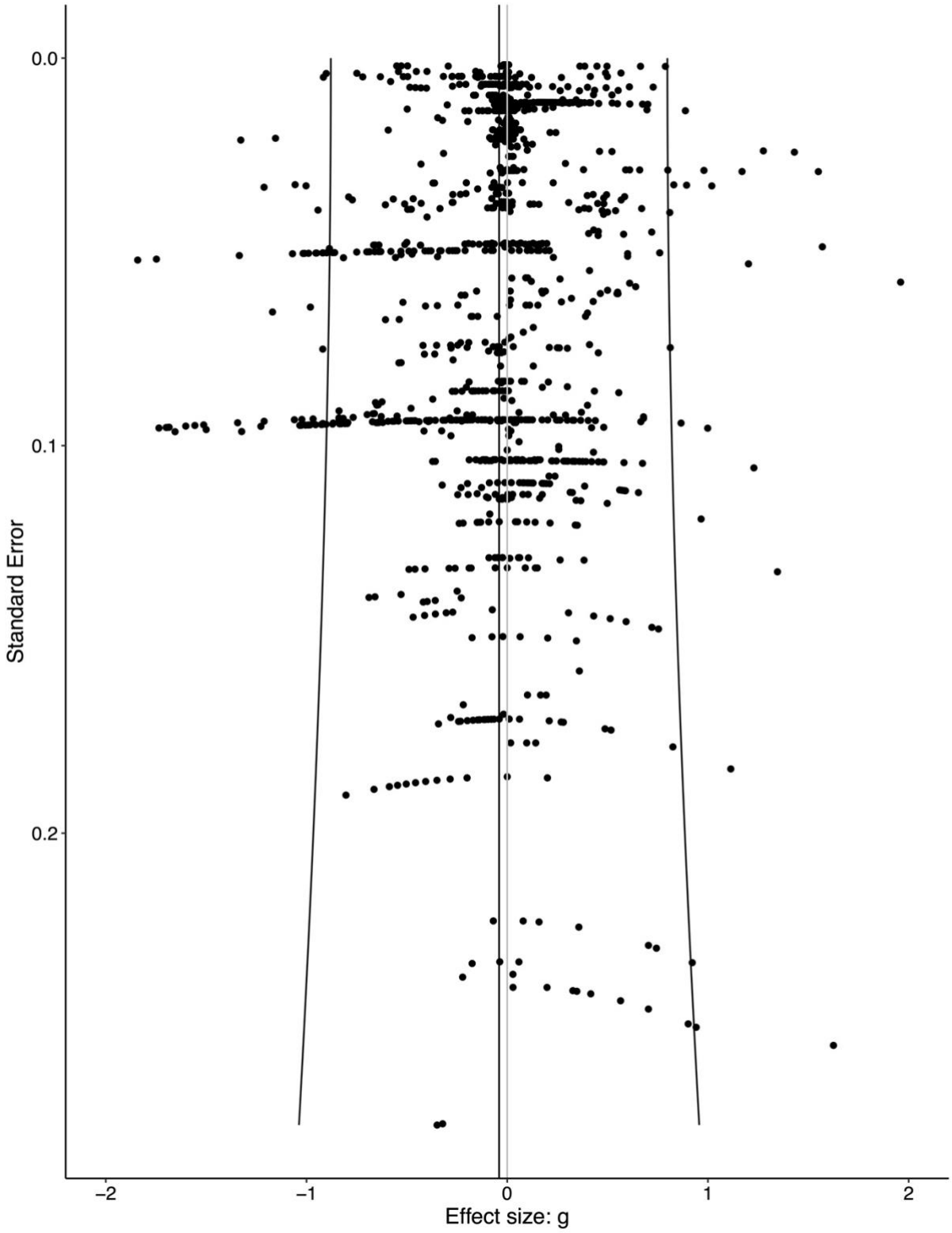
Forest Plot of the Meta-Analysis



Note. J = number of effects in each study

Appendix A3

Funnel Plot of Standard Error by Effect Size



APPENDIX B: Supporting information for Chapter 4

Appendix B1

Multilevel Unstandardized Parameter Estimates for Reading Self-Concept

	Model 0	Model 1	Model 2	Model 3
Intercept	-0.48 (0.02)***	-0.11 (0.08)	-0.11 (0.08)	-0.10 (0.08)
Student level				
Retention		0.03 (0.02)	0.04 (0.03)	0.01 (0.03)
Age		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Male		-0.06 (0.02)**	-0.06 (0.02)*	-0.06 (0.02)*
ESCS[a]		0.07 (0.01)***	0.07 (0.01)***	0.08 (0.01)***
Immigration status		0.02 (0.04)	0.02 (0.04)	0.03 (0.04)
Home language		-0.24 (0.06)***	-0.24 (0.06)***	-0.26 (0.06)***
HEDRES[b]		0.10 (0.01)***	0.10 (0.01)***	0.10 (0.01)***
EMOSUPP[c]		0.06 (0.01)***	0.06 (0.01)***	0.07 (0.01)***
Reading score		0.36 (0.03)***	0.36 (0.03)***	0.37 (0.03)***
Math score		-0.31 (0.03)***	-0.31 (0.03)***	-0.31 (0.03)***
Science score		0.12 (0.03)***	0.13 (0.03)***	0.12 (0.03)***
School level				
Retention composition			-0.01 (0.02)	-0.07 (0.03)*
Public school		0.10 (0.07)	0.11 (0.07)	0.11 (0.07)
School ESCS[a] composition		-0.03 (0.02)	-0.04 (0.03)	-0.04 (0.03)
School immigrant composition		-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Retention X Retention composition				0.08 (0.02)***
Between school variance	0.09 (0.29)	0.08 (0.28)	0.08 (0.28)	0.08 (0.28)
Within school variance	0.34 (0.59)	0.31 (0.56)	0.31 (0.56)	0.31 (0.56)

Note. Standard errors are in parenthesis. [a]Index of economic, social, and cultural background; [b]Home educational resources; [c]Parental emotional support. Standard errors are in parenthesis

*p < .05, ** p < .01, ***p < .001

Appendix B2

Multilevel Unstandardized Parameter Estimates for Task Orientation

	Model 0	Model 1	Model 2	Model 3
Intercept	-0.18 (0.03)***	0.11 (0.10)	0.11 (0.10)	0.12 (0.10)
Student level				
Retention		-0.12 (0.03)***	-0.11 (0.03)***	-0.16 (0.03)***
Age		0.03 (0.01)	0.02 (0.01)	0.02 (0.01)
Male		-0.22 (0.03)***	-0.22 (0.03)***	-0.21 (0.03)***
ESCS[a]		0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
Immigration status		0.18 (0.05)**	0.18 (0.05)***	0.20 (0.05)***
Home language		0.07 (0.07)	0.07 (0.07)	0.05 (0.07)
HEDRES[b]		0.16 (0.01)***	0.16 (0.01)***	0.16 (0.01)***
EMOSUPP[c]		0.20 (0.01)***	0.20 (0.01)***	0.21 (0.01)***
Reading score		-0.12 (0.04)***	-0.11 (0.04)***	-0.10 (0.04)**
Math score		-0.06 (0.04)	-0.06 (0.04)	-0.07 (0.04)
Science score		0.20 (0.04)***	0.20 (0.04)***	0.20 (0.04)***
School level				
Retention composition			-0.04 (0.02)	-0.13 (0.04)***
Public school		-0.09 (0.08)	-0.08 (0.08)	-0.07 (0.08)
School ESCS[a] composition		0.01 (0.03)	-0.03 (0.03)	-0.04 (0.03)
School immigrant composition		-0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
Retention X Retention composition				0.13 (0.03)***
Between school variance	0.12 (0.35)	0.10 (0.32)	0.10 (0.32)	0.10 (0.32)
Within school variance	0.52 (0.72)	0.47 (0.68)	0.47 (0.68)	0.47 (0.68)

Note. Standard errors are in parenthesis. [a]Index of economic, social, and cultural background; [b]Home educational resources; [c]Parental emotional support. Standard errors are in parenthesis

*$p < .05$, ** $p < .01$, ***$p < .001$

Appendix B3

Multilevel Unstandardized Parameter Estimates for Self-Enhancing Orientation

	Model 0	Model 1	Model 2	Model 3
Intercept	-0.03 (0.03)	-0.06 (0.10)	-0.06 (0.10)	-0.05 (0.10)
Student level				
Retention		-0.05 (0.03)	-0.05 (0.03)	-0.07 (0.03)
Age		0.02 (0.01)	0.02 (0.01)	0.01 (0.01)
Male		0.42 (0.03)***	0.42 (0.03)***	0.42 (0.03)***
ESCS[a]		0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
Immigration status		-0.14 (0.06)*	-0.14 (0.06)*	-0.14 (0.06)*
Home language		-0.16 (0.07)*	-0.16 (0.07)*	-0.17 (0.07)*
HEDRES[b]		0.04 (0.02)*	0.04 (0.02)*	0.04 (0.02)*
EMOSUPP[c]		0.11 (0.01)***	0.11 (0.01)***	0.11 (0.01)***
Reading score		-0.15 (0.04)***	-0.15 (0.04)***	-0.14 (0.04)***
Math score		-0.21 (0.04)***	-0.21 (0.04)***	-0.21 (0.04)***
Science score		0.34 (0.04)***	0.34 (0.04)***	0.34 (0.04)***
School level				
Retention composition			-0.01 (0.03)	-0.01 (0.03)
Public school		-0.01 (0.08)	-0.01 (0.08)	-0.01 (0.08)
School ESCS[a] composition		-0.01 (0.03)	-0.01 (0.03)	-0.01 (0.03)
School immigrant composition		-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Retention X Retention composition				0.03 (0.03)
Between school variance	0.11 (0.33)	0.10 (0.31)	0.10 (0.31)	0.10 (0.31)
Within school variance	0.54 (0.74)	0.51 (0.72)	0.51 (0.72)	0.51 (0.72)

Note. Standard errors are in parenthesis. [a]Index of economic, social, and cultural background; [b]Home educational resources; [c]Parental emotional support. Standard errors are in parenthesis

*p < .05, ** p < .01, ***p < .001

Appendix B4

Multilevel Unstandardized Parameter Estimates for School Belonging

	Model 0	Model 1	Model 2	Model 3
Intercept	0.08 (0.03)**	0.44 (0.10)***	0.44 (0.10)***	0.45 (0.10)***
Student level				
Retention		-0.09 (0.03)**	-0.09 (0.03)**	-0.12 (0.03)***
Age		-0.03 (0.01)*	-0.03 (0.01)*	-0.03 (0.01)**
Male		0.20 (0.03)***	0.20 (0.03)***	0.20 (0.03)***
ESCS[a]		0.06 (0.02)***	0.06 (0.02)***	0.06 (0.02)***
Immigration status		-0.19 (0.05)***	-0.19 (0.05)***	-0.18 (0.05)**
Home language		0.03 (0.07)	0.03 (0.07)	0.01 (0.07)
HEDRES[b]		0.09 (0.01)***	0.09 (0.01)***	0.09 (0.01)***
EMOSUPP[c]		0.22 (0.01)***	0.22 (0.01)***	0.23 (0.01)***
Reading score		0.09 (0.03)**	0.08 (0.03)*	0.10 (0.03)**
Math score		-0.30 (0.03)***	-0.30 (0.03)***	-0.30 (0.03)***
Science score		0.29 (0.04)***	0.29 (0.04)***	0.29 (0.04)***
School level				
Retention composition			0.01 (0.03)	-0.05 (0.03)
Public school		-0.32 (0.07)***	-0.32 (0.07)***	-0.32 (0.08)***
School ESCS[a] composition		0.02 (0.03)	0.03 (0.03)	0.02 (0.03)
School immigrant composition		0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
Retention X Retention composition				0.09 (0.03)**
Between school variance	0.14 (0.38)	0.10 (0.32)	0.10 (0.32)	0.10 (0.32)
Within school variance	0.47 (0.69)	0.43 (0.65)	0.43 (0.65)	0.43 (0.65)

Note. Standard errors are in parenthesis. [a]Index of economic, social, and cultural background; [b]Home educational resources; [c]Parental emotional support. Standard errors are in parenthesis

*p < .05, ** p < .01, ***p < .001

Appendix B5

Multilevel Unstandardized Parameter Estimates for School Utility Value

	Model 0	Model 1	Model 2	Model 3
Intercept	0.20 (0.03)***	0.47 (0.10)***	0.47 (0.10)***	0.49 (0.10)***
Student level				
Retention		-0.03 (0.03)	-0.03 (0.03)	-0.08 (0.03)*
Age		-0.06 (0.01)***	-0.06 (0.01)***	-0.06 (0.01)***
Male		-0.16 (0.03)***	-0.16 (0.03)***	-0.16 (0.03)***
ESCS[a]		0.02 (0.02)	0.02 (0.02)	0.02 (0.02)
Immigration status		0.07 (0.05)	0.07 (0.05)	0.09 (0.05)
Home language		0.08 (0.07)	0.08 (0.07)	0.06 (0.07)
HEDRES[b]		0.11 (0.01)***	0.11 (0.01)***	0.11 (0.01)***
EMOSUPP[c]		0.22 (0.01)***	0.22 (0.01)***	0.22 (0.01)***
Reading score		0.22 (0.03)***	0.22 (0.03)***	0.24 (0.03)***
Math score		0.03 (0.03)	0.03 (0.03)	0.03 (0.03)
Science score		-0.10 (0.04)*	-0.10 (0.04)**	-0.10 (0.04)**
School level				
Retention composition			-0.03 (0.03)	-0.12 (0.03)***
Public school		-0.05 (0.07)	-0.04 (0.07)	-0.03 (0.07)
School ESCS[a] composition		0.04 (0.03)	0.02 (0.03)	0.01 (0.03)
School immigrant composition		0.01 (0.02)	0.01 (0.02)	0.01 (0.02)
Retention X Retention composition				0.14 (0.03)***
Between school variance	0.11 (0.33)	0.09 (0.30)	0.09 (0.30)	0.09 (0.30)
Within school variance	0.49 (0.70)	0.43 (0.66)	0.43 (0.66)	0.43 (0.66)

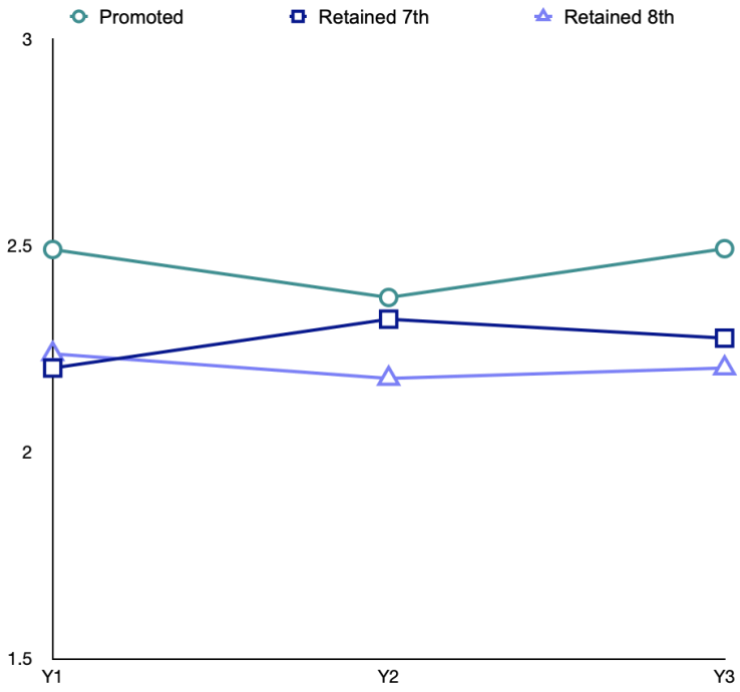
Note. Standard errors are in parenthesis. [a]Index of economic, social, and cultural background; [b]Home educational resources; [c]Parental emotional support. Standard errors are in parenthesis

*$p < .05$, ** $p < .01$, ***$p < .001$

APPENDIX C. Supporting information for Chapter 6

Appendix C1

Math Achievement Development Across Three School Years of Promoted Students, Retained Students in 7th Grade, and Retained Students in 8th Grade



Appendix C2

Portuguese Achievement Development Across Three School Years of Promoted Students, Retained Students in 7th Grade and Retained Students in 8th Grade.